UNIVERSITÉ PARIS-DAUPHINE

ÉCOLE DOCTORALE DE DAUPHINE

CENTRE DE RECHERCHE EN MATHÉMATIQUES DE LA DÉCISION

THÈSE

pour l'obtention du titre de

DOCTEUR EN SCIENCES, DISCIPLINE MATHÉMATIQUES APPLIQUÉES

présentée et soutenue par

Yang LU

le 24/06/2015

Analyse de survie bivariée à facteurs latents : Théorie et applications à la mortalité et à la dépendance

Bivariate Survival Analysis with Latent Factors : Theory and Applications to Mortality and Long-Term Care

Directeur de thèse

Prof. Christian GOURIÉROUX (Université Paris-Dauphine)

Rapporteurs

Prof. Michel DENUIT (Université catholique de Louvain)

Prof. Christian GENEST (McGill University)

Jury

Prof. Serge DAROLLES (Université Paris-Dauphine)

Prof. Michel DENUIT (Université catholique de Louvain)

Dr. Xavier D'HAULTFOEUILLE (CREST)

Prof. Christian GOURIÉROUX (Université Paris-Dauphine)

Prof. Armelle GUILLOU (Université de Strasbourg)

Prof. Donatien HAINAUT (ESC Rennes School of Business)

Remerciements

Mes remerciements vont en premier lieu à mon directeur de thèse Christian Gouriéroux, tant pour sa disponibilité que pour ses nombreux conseils. Sa rigueur, sa connaissance scientifique et ses conseils pour le job market m'ont été précieux.

Je remercie SCOR Global Life, qui a financé cette thèse, fourni des données précieuses, et m'a permis de participer à de nombreuses conférences internationales. Des discussions avec Daria, Tiziana, Razvan, Thibaut, Julien, Guillaume et Vincent ont été très utiles.

Ma profonde reconnaissance va aussi aux membres du laboratoire de Finance et Assurance du CREST : Jean-Michel ZAKOIAN, Christian FRANCQ, Bertrand VILLENEUVE, Olivier WINTENBERGER, Serge DAROLLES, Christian ROBERT, Xavier MILHAUD, Olivier LOPEZ et Jean-David FERMANIAN pour des discussions intéressantes.

Je souhaite remercier chaleureusement mes deux rapporteurs : Michel DENUIT et Christian GENEST, ainsi que les autres membres du jury : Donatien HAINAUT, Xavier D'HAULTFOEUILLE, Serge DAROLLES et Armelle GUILLOU.

Je pense également à mes amis, doctorants, anciens doctorants, ou non doctorants : les Guillaumes, Erwan, Jean-Cyprien, Jérémy, Yiyi, Xiaofei, Nicolas, Benjamin, Dongli, Jianfeng, etc.

Evidemment, un grand merci à toute ma famille : mes parents, ma femme Shize et mes beaux parents.

Résumé

Cette thèse étudie quelques problèmes d'identification et d'estimation dans les modèles de survie bivariée, avec présence d'hétérogénéités individuelles et facteurs communs stochastiques.

Le Chapitre I introduit le cadre général.

Le Chapitre II propose un modèle pour la mortalité des deux époux dans un couple. Il permet de distinguer deux types de dépendance : l'effet de deuil et l'effet lié au facteur de risque commun des deux époux. Une analyse de leurs effets respectifs sur les primes d'assurance écrites sur deux têtes est proposée.

Le Chapitre III montre que, sous certaines hypothèses raisonnables, on peut identifier l'évolution jointe du risque d'entrer en dépendance et du risque de mortalité, à partir des données de mortalité par cohortes. Une application à la population française est proposée.

Le Chapitre IV étudie la queue de distribution dans les modèles de survie bivariée. Sous certaines hypothèses, la loi jointe des deux durées résiduelles converge, après une normalization adéquate. Cela peut être utilisé pour analyser le risque parmi les survivants aux âges élevés. Parallèlement, la distribution d'hétérogénéité parmi les survivants converge vers une distribution semi-paramétrique.

Mots clés : facteurs latents (statiques ou dynamiques), risques concurrents, effet de traitement, valeurs extrêmes, identification non-paramétrique, mortalité, dépendance des personnes âgées, risque de longévité, assurance-vie.

Abstract

This thesis comprises three essays on identification and estimation problems in bivariate survival models with individual and common frailties.

The first essay proposes a model to capture the mortality dependence of the two spouses in a couple. It allows to disentangle two types of dependencies : the broken heart syndrome and the dependence induced by common risk factors. An analysis of their respective effects on joint insurance premia is also proposed.

The second essay shows that, under reasonable model specifications that take into account the longevity effect, we can identify the joint distribution of the long-term care and mortality risks from the observation of cohort mortality data only. A numerical application to the French population data is proposed.

The third essay conducts an analysis of the tail of the joint distribution for general bivariate survival models with proportional frailty. We show that, under appropriate assumptions, the distribution of the joint residual lifetimes converges to a limit distribution, up to a normalization. This can be used to analyze the mortality and long-term care risks at advanced ages. In parallel, the heterogeneity distribution among survivors converges also to a semi-parametric limit distribution. Properties of the limit distributions, their identifiability from the data, as well as their implications are discussed.

Keywords : Static and dynamic latent factors, competing risks, treatment effects, extreme values, non-parametric identification, mortality, longevity risk, life insurance.

Table des matières

Chapitre I

Introduction

Le phénomène de la longévité humaine pose de plus en plus de nouveaux défis pour notre société. Cette thèse a pour but d'étudier quelques modèles de survie bivariée avec des facteurs latents, qui seront appliqués à la prévision des risques de mortalité et de dépendance des personnes âgées.

Les données de survie bivariée interviennent quand plusieurs variables de durées sont potentiellement observables. Cela couvre trois cas principaux. Dans le premier cas, les deux variables de durée sont observables. Cela est par exemple le cas quand nous étudions les durées de vie des deux époux dans un couple. Le deuxième cas est dit semi-concurrent, dans le sens où l'une des variables est latente, c'est-à-dire n'est observable que dans certains cas. Par exemple, l'âge d'entrée en dépendance d'un individu est observable uniquement si cet individu entre réellement en dépendance au cours de sa vie. Enfin, dans un modèle à risques concurrents, on observe seulement une variable de durée, c'est-à-dire la plus petite d'entre elles, et la cause de décès, c'est-à-dire l'indice de la variable réellement observée.

Les données de survie étant des données individuelles, il est naturel de tenir compte de la présence d'hétérogénéité (observée ou non observée) des individus. Les premiers modèles à facteur d'hétérogénéité latente sont dus à Vaupel et al. (1979) et Lancaster (1979). L'hétérogénéité y a été introduite avec un effet proportionnel sur l'intensité. L'introduction de l'hétérogénéité permet de contrôler le biais de dépendance négative du au processus de sélection dans une population hétérogène. Intuitivement, pour des caractéristiques observables identiques, les individus les plus risqués (le risque étant mesuré par l'hétérogénéité non observée) décèdent plus vite, ce qui entraîne une sélection endogène au sein de la population. Dans les modèles de survie à deux

variables, deux facteurs d'hétérogénéité peuvent être introduits et ces facteurs d'hétérogénéité ont un effet supplémentaire, qui est de contrôler la dépendance entre les deux variables de durée.

Le risque de longévité est le risque que les individus vivent, en moyenne, plus longtemps que prévu. Dans le passé, les prévisions d'espérance de vie ont toujours sous-estimé la véritable amélioration. L'allongement de la durée de vie conduit également à la hausse du coût de la dépendance des personnes âgées. La longévité est un effet incertain, c'est-à-dire stochastique, et en général elle a des effets sur toutes les variables de durées liées à la vie humaine, par exemple l'âge d'entrée en dépendance, la durée passée en dépendance, l'âge de décès des deux époux, ou même les causes de décès. Par conséquent, dans tous les modèles bivariés considérés dans cette thèse, on introduit un facteur latent stochastique de longévité, qui est commun pour tous les risques étudiés.

Cette thèse contribue à la littérature sur la survie bivariée en proposant des spécifications adaptées aux problèmes rencontrés dans les domaines de l'assurance-vie et de l'assurance dépendance. Pour chacun des problèmes de survie abordés, nous allons suivre la démarche méthodologique ci-dessous :

1. Spécification du modèle : l'étude de ses propriétés théoriques, notamment des conséquences sur les variables observées des diverses hypothèses.
2. Identification du modèle : les modèles de durée que nous considérons sont soumis à des problèmes d'observabilité : il peut s'agir de variables non observables comme les hétérogénéités latentes, ou partiellement observables, lorsque les durées sont soumises à des censures, comme dans le cas des risques concurrents. Une conséquence de ces problèmes de non observabilité est la non identifiabilité potentielle de certaines paramètres (fonctionnels) des modèles. Nous étudions de façon systématique cette question de l'identifiabilité.
3. Estimation du modèle : une fois défini un modèle adapté à l'application et identifiable, nous proposons une méthode d'estimation ou de valorisation, et l'appliquerons ensuite aux données réelles. Dans le cas où le modèle peut être utilisé pour la valorisation (par exemple dans le chapitre II), nous étudions également les implications du modèle en terme de primes d'assurance.

Dans ce chapitre introductif, nous commençons par décrire le risque de longévité rencontré en assurance-vie. Nous rappelons ensuite les modèles de base utilisés pour prévoir la mortalité future, c'est-à-dire les modèles de type Lee, Carter (1992). Dans un troisième temps, nous replaçons ces

modèles de mortalité de base dans le contexte plus général des modèles de survie, introduisons les modèles de survie bivariée et donnons des exemples d'applications de ces modèles bivariés en assurance. Enfin, nous expliquons l'importance de tenir compte des facteurs latents, qui peuvent être individuels ou communs, statiques ou dynamiques. Nous donnons également à la fin du chapitre les résumés des trois articles écrits pendant la thèse, qui servent de base aux trois chapitres suivants.

I.1 Le risque de longévité

Le risque de longévité est le risque incertain de diminution des taux de mortalité. Il est nécessaire d'insister d'emblée sur le fait que non seulement les espérances de vie ont une tendance haussière, mais aussi que cette augmentation est stochastique, car l'évolution de la mortalité future est incertaine.

Lors de ces dernières années, la durée de la vie humaine n'a cessé d'augmenter et entraîne un vieillissement de la population. Cet allongement de la durée de vie humaine s'explique principalement par les progrès de la médecine et l'amélioration des conditions de vie. Historiquement son impact sur le système de retraite, qu'il soit public ou privé, a été généralement sous-estimé, et il est devenu de plus en plus urgent, pour les organismes assureurs et les systèmes de retraite, de prédire de façon fiable l'évolution de la mortalité dans le futur.

Du point de vue d'un (ré)assureur, le risque de longévité correspond au risque que la population assurée vive plus longtemps que la prévision faite à partir de la table de mortalité utilisée. Il présente plusieurs caractéristiques : premièrement c'est un phénomène en constante évolution, et difficile à prévoir car influencé par divers facteurs tels que les politiques budgétaires, les progrès de la médecine, l'évolution des modes de vie (e.g. fumeur/non fumeur). Deuxièmement c'est un risque non mutualisable du fait de ces facteurs extérieurs communs. Troisièmement, c'est un risque à très long terme, de l'ordre de plusieurs dizaines d'années, et les écarts de tendances d'évolutions sont difficiles à détecter.

I.1.1 Quelques produits financiers sensibles au risque de longévité

Décrivons quelques contrats sensibles au risque de longévité.

i). Une **rente** est un contrat d'assurance dans lequel l'assureur s'engage à verser une série annuelle de paiements à l'acheteur, contre une prime d'assurance, payée au moment de la

souscription du contrat. La figure I-1 fournit le schéma de flux d'un contrat de rente type.

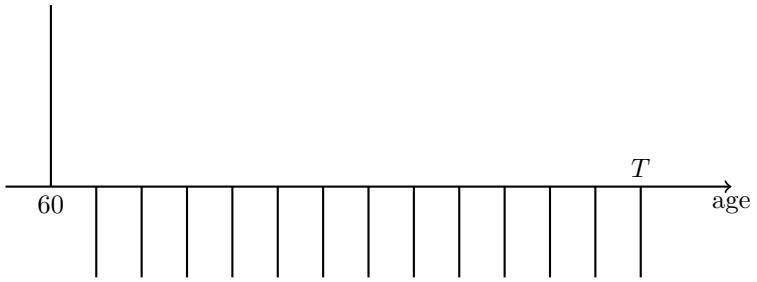


FIGURE I-1: Flux financier d'un contrat de rente, souscrit par un individu à l'âge de 60 ans. Il paie une seule prime d'assurance à la souscription du contrat, et reçoit, à partir de l'année suivante, un paiement annuel de la part de l'assureur, jusqu'à l'âge de décès T, qui est stochastique et supérieur à 60. Dans cette figure, le paiement annuel de l'assureur a été supposé, pour des raisons de simplicité, constant. Il peut exister des cas où le paiement annuel est croissant, à un taux fixé, ou à un taux variable indexé sur l'inflation.

ii). D'autres institutions financières, largement concernées par le risque de longévité, sont les **fonds de pension**. Dans de nombreux pays tels que le Royaume-Uni, les Etats-Unis ou les Pays-Bas , les prestations de fonds de pension que versent les employeurs à leurs anciens employés constituent la source principale du revenu après la retraite.

Par exemple un fonds de pension à prestations définies est un fonds financé par un employeur pour gérer la retraite de ses employés. L'employeur s'engage à verser une somme annuelle prédéfinie au moment du départ en retraite de l'employé, et ce jusqu'à sa mort. Par conséquent, un fonds de pension a la même obligation financière qu'un assureur ayant vendu un contrat de rente.

iii). Il existe aussi d'autres formes de retraite comme des **retraites par répartition**, généralement gérées soit par des organismes publics, soit par des caisses de retraite professionnelles. Le facteur longévité influe directement sur l'évolution de la structure par âge de la population assurée par la caisse et donc sur la répartition entre actifs et retraités.

Ces trois exemples (contrat de rente, fonds de pension, retraite par répartition) sont les principales sources de financement de retraite des individus (avec aussi, bien sûr, l'épargne individuelle). On renvoie le lecteur intéressé au rapport de Gruber and Wise (1999) pour une comparaison entre différents pays des répartitions du financement de la retraite entre ces différents moyens.

Il existe aussi d'autres types de contrats d'assurance où le risque de longévité est présent avec d'autres risques biométriques. Un exemple typique est celui de l'assurance dépendance.

Une personne entre en dépendance quand elle perd une certaine autonomie, mesurée par

l'incapacité d'accomplir sans assistance des actes ordinaires de la vie quotidienne tels que prendre ses repas, faire sa toilette, se déplacer, s'habiller.

Dans un **contrat d'assurance dépendance**, le client paye une prime régulière à l'assureur en échange de la garantie d'une rente s'il entre en dépendance. Dans le cas où le client n'entre pas en dépendance durant la vie du contrat, l'assureur n'a pas de sinistres à payer. La Figure I-2 fournit un schéma illustratif des flux financiers dans le cas où il décède (à l'âge T_2) sans passer par une perte d'autonomie. Dans le cas où le client entre en dépendance à l'âge T_1 avant de décéder à l'âge T_2, avec $T_1 < T_2$, les flux financiers sont représentés par la Figure I-3.

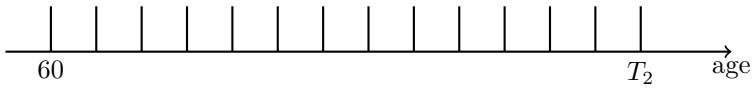


FIGURE I-2: Flux dans le cas où l'individu décède directement sans passer par la phase de dépendance. Dans ce cas, les flux financiers se limitent aux seules cotisations payées par l'assuré entre l'âge 60 et la date de décès T_2.

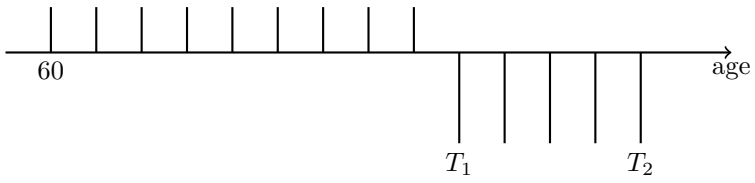


FIGURE I-3: Flux financier d'un contrat d'assurance de dépendance souscrit à 60 ans. dans le cas où l'individu entre en dépendance à l'âge T_1, et décède à l'âge T_2. Les flux financiers incluent non seulement les cotisations payées par l'assuré entre l'âge 60 et T_1, mais également les sinistres payés par l'assureur effectués entre T_1 and T_2.

Un assureur fournissant des contrats de dépendance est concerné par plusieurs types de risque de longévité. Premièrement, les taux de mortalité des personnes sans perte d'autonomie est en baisse dans le temps, à âge donné, ce qui se traduit par un accroissement de la population à risque. Deuxièmement, le taux de mortalité des personnes dépendantes diminue à âge donné. Ceci pourrait induire une durée moyenne de séjour de plus en plus longue dans l'état de dépendance.

Il y a une littérature actuarielle abondante sur la gestion et la tarification du risque de longévité pour les contrats de rente pour les compagnies d'assurance [voir par exemple Wills and Sherris (2010); Bauer et al. (2010); Li and Hardy (2011)], ou pour les produits d'assurance dépendance [voir Levantesi and Menzietti (2012)]. Ces points ne sont pas abordés dans cette thèse et nous renvoyons le lecteur intéressé aux papiers cités.

I.1.2 Données disponibles

Il existe évidemment des bases de données internes aux compagnies d'assurance concernant leur propre clientèle d'assurés. Ces bases sont souvent assez hétérogènes et soumises à des restrictions de disponibilité.

Dans nos applications nous utilisons une base gérée par l'Université de Californie, Berkeley, qui est la *Human Mortality Database*[1]. Cette base de mortalité présente l'avantage de concerner l'ensemble des populations de divers pays industrialisés (évitant de ce fait des biais de représentativité), d'être bien renseignée, et maintenue à jour régulièrement. De plus elle est en accès libre. Ceci facilitera donc la comparaison de nos résultats avec ceux d'autres études parallèles. On dispose en général, pour chaque individu de la population étudiée, de sa date de naissance et de sa date de mort, ainsi que d'autres caractéristiques telles que son sexe ou sa cause de décès.

La différence entre des données relatives aux populations nationales et celles des clientèles d'assurés est que, dans le premier cas, l'historique d'observation est beaucoup plus long, les taux de mortalité sont moins volatils que pour les populations d'assurés, du fait de la plus grande taille de la population nationale. Cependant les bases de données des assureurs peuvent être plus renseignées sur certaines caractéristiques individuelles, notamment financières, et une segmentation plus fine est souvent possible en tenant compte des aspects fumeur/non fumeur, des revenus, ou d'autres investissements financiers que l'assurance-vie.

I.2 Le modèle de Lee-Carter

Dans ce paragraphe nous rappelons le principal modèle de mortalité incluant un facteur stochastique de longévité. Il s'appuie sur une modélisation du taux de mortalité indexé par l'âge et le temps calendaire. Nous utilisons pour le décrire les notations actuarielles standard, même si dans les chapitres de la thèse les notations probabilistes classiques sont utilisées. Ainsi le taux de mortalité à la date t pour un individu d'âge x, c'est-à-dire né à la date $c = t - x$, est noté $q_x(t)$. Il est défini par :

$$q_x(t) = \mathbb{P}[T = x | T \geq x, c = t - x].$$

1. www.mortality.org

I.2.1 Le modèle de base

Le modèle introduit par Lee, Carter (1992) est le premier modèle de mortalité stochastique connu. Il est devenu, depuis sa publication, un standard pour les actuaires et un point de départ pour introduire d'autres modèles plus sophistiqués. Ce modèle suppose que l'évolution dans le temps des taux de mortalité est entraînée par un facteur commun, avec des degrés de sensibilité différents pour des âges différents. Lee et Carter proposent la modélisation suivante :

$$\ln q_x(t) = \alpha_x + \beta_x \kappa_t + \epsilon_{x,t}, \tag{I-1}$$

où les termes d'erreur $\epsilon_{x,t}$ sont centrés, indépendants et de même variance σ^2 (homoscédasticité conditionnelle).

Donnons ici la signification de chaque paramètre.

— α_x est un paramètre de niveau des taux instantanés de mortalité à l'échelle logarithmique.

— κ_t est le facteur inobservable, stochastique, qui entraîne l'évolution de toutes les séries temporelles $(q_x(t)), t_{\min} \leq t \leq t_{\max}$ dans le temps.

— β_x décrit la sensibilité de $\log q_x(t)$ par rapport au facteur κ_t.

Pour rendre le modèle identifiable, Lee, Carter proposent d'ajouter deux contraintes supplémentaires. La première porte sur les sensibilités au facteur de longévité :

$$\sum_{x=x_{\min}}^{x_{\max}} \beta_x = 1, \tag{I-2}$$

la seconde sur le niveau moyen du facteur commun :

$$\mathbb{E}[\kappa_t] = 0. \tag{I-3}$$

Pour estimer les paramètres α_x, β_x et filtrer les valeurs du facteur latent, Lee, Carter proposent d'employer une approche par moindres carrés asymptotiques dans laquelle les valeurs (stochastiques) κ_t sont considérés comme des paramètres additionnels, et la contrainte identifiante I.3 est remplacée par sa contrepartie empirique. Notons $\hat{q}_x(t)$ les taux de mortalité observés proches des $q_x(t)$, si le nombre de survivants d'âge x à la date t est suffisamment grand.

Les approximations des paramètres sous-jacents et des valeurs du facteur latent sont obtenues en résolvant le problème de moindres carrés ordinaires contraint :

$$(\alpha, \beta, \kappa) = \arg\min_{\alpha,\beta,\kappa} \sum_{x=x_{\min}}^{x_{\max}} \sum_{t=t_{\min}}^{t_{\max}} \Big(\ln \hat{q}_x(t) - \alpha_x - \beta_x \kappa_t \Big)^2, \tag{I-4}$$

sous les contraintes identifiantes (approchée pour la seconde) : $\sum_{x=x_{\min}}^{x_{\max}} \beta_x = 1$, $\sum_{t=t_{\min}}^{t_{\max}} \kappa_t = 0$.

Cette approche pragmatique permet non seulement de connaître des approximations des coefficients α_x et β_x, mais aussi de reconstituer une trajectoire (filtrée) du facteur, qui peut être utilisée pour étudier sa dynamique. En effet, sans connaissance de cette dynamique, il n'y a aucune possibilité d'utiliser le modèle dans un but de prévision.

I.2.2 Limites et premières extensions du modèle de Lee-Carter

Du fait de leurs simplicités, le modèle de base et la méthode d'estimation souffrent de certaines limites. Ceci a conduit à beaucoup de variantes et d'extensions.

i) **Méthode d'estimation** La méthode initiale consistant à remplacer directement les taux de mortalité théoriques par les taux observés dans l'équation (I-1) et à appliquer les moindres carrés ordinaires ne prend pas en compte les erreurs d'observation. Plus précisément, on a :

$$\ln \hat{q}_{x,t} = \alpha_x + \beta_x \kappa_t + \epsilon_{x,t} + \eta_{x,t},$$

avec $\eta_{x,t} = \ln \hat{q}_{x,t} - \ln q_{x,t}$. Ces erreurs de mesure peuvent introduire de l'hétéroscédasticité conditionnelle et des aspects non gaussiens. Ainsi, aux grands âges, le nombre d'individus dans la population à risque est très réduit. Or, le risque de longévité est particulièrement important aux grands âges. Ceci a deux effets : la précision sur $q_{x,t}$ est plus faible et on ne peut pas considérer $\hat{q}_{x,t}$ comme une approximation gaussienne de $q_{x,t}$. Dans ce cas, une hypothèse Poissonnienne est plus adaptée (voir, e.g. Brouhns et al. (2002)) :

$$d_x(t) \sim \text{Poisson}(e_x(t) \exp(\alpha_x + \beta_x \kappa_t)).$$

L'estimation du modèle de Poisson ci-dessus passe soit par une méthode de maximum de vraisemblance, plus efficace, soit par des techniques bayésiennes (voir e.g. Czado et al. (2005), lorsque $\epsilon_{x,t} = 0$).

ii) **Spécification de la dynamique du facteur** La méthode d'estimation de paramètres proposée par Lee, Carter (1992) est de type nonparamétrique au sens où aucune hypothèse n'est faite concernant l'évolution du facteur κ_t (ou la forme des coefficients de sensibilité). Il en découle des valeurs estimées des κ_t très erratiques dans le temps, bien qu'elles présentent une tendance globale baissière, et sensibles au choix de la période d'observation, notamment lors d'une mise à jour. En effet, la condition d'identification approchée des κ_t dépend de la période d'observation retenue.

Il est donc apparu utile de modifier le modèle, en ajoutant des hypothèses dynamiques et paramétriques sur κ_t. Cela peut être un modèle type autorégressif moyenne mobile intégré (ARIMA) [voir e.g. Lee, Carter (1992)] ou plus généralement un processus admettant une écriture sous forme espace-état. Par exemple, on pourra écrire [voir Pedroza (2006)] :

$$\ln q_x(t) = \alpha_x + \beta_x \kappa_t + \epsilon_{x,t},$$

$$\kappa_t = \kappa_{t-1} + \theta + \omega_t, \tag{I-5}$$

avec les ω_t indépendants des $\epsilon_{x,t}$ et i.i.d. gaussiens centrés. La dynamique de marche aléatoire avec effet de translation du processus (κ_t) permet l'introduction à la fois de tendances déterministes et stochastiques.

Cette spécification paramétrique permet alors d'estimer de façon asymptotiquement efficace les paramètres du modèle ($\alpha_x, \beta_x, \theta$, et les variances de $\epsilon_{x,t}$ et $\omega_{x,t}$) par maximum de vraisemblance. Ainsi l'estimation se fait en une seule étape au lieu de deux étapes dans l'approche initiale. De plus les valeurs inconnues du facteur peuvent aussi être filtrés de façon optimale par l'utilisation du filtre de Kalman.

iii). **Ajout de facteurs temporels.** Le modèle de Lee-Carter peut sous-estimer la vitesse d'amélioration de la durée de vie à long terme. Ceci est notamment dû au fait que le modèle de base inclut un seul facteur temporel. Ainsi, après une phase de forte diminution de mortalité aux jeunes âges jusqu'aux années 50, l'amélioration de la survie est devenue de plus en plus marquée pour les personnes âgées : ce sont elles qui contribuent le plus actuellement à la hausse de l'espérance de vie. Cette observation conduit à un modèle du type [voir Renshaw and Haberman (2003)] :

$$\ln q_x(t) = \alpha_x + \beta_{1,x} \kappa_{1,t} + \beta_{2,x} \kappa_{2,t} + \epsilon_{x,t},$$

où le second facteur traduit l'accélération de l'amélioration aux grands âges. Un cas particulier est le modèle de Cairns et al. (2006) (appelé aussi C.B.D.) dans lequel des hypothèses sont aussi faites sur la forme des coefficients de sensibilité $\beta_{1,x}, \beta_{2,x}$ et α_x :

$$\text{logit } \hat{q}(t,x) = \kappa_t^{(1)} + \kappa_t^{(2)}(x - \overline{x}) + \epsilon_{t,x} \tag{I-6}$$

Leurs travaux ont abouti à plusieurs variantes. Une comparaison empirique a été notamment proposée par Cairns et al. (2009). Cette extension peut aussi être conduite en introduisant des hypothèses dynamiques sur les lois jointes des deux facteurs $\kappa_t^{(1)}, \kappa_t^{(2)}$ caractérisant le phénomène de longévité.

iv) **Prise en compte de l'effet cohorte.** En démographie, l'effet cohorte correspond à une ou plusieurs générations dont l'amélioration de mortalité est particulièrement importante par rapport aux générations voisines. Ceci est par exemple le cas pour la *génération dorée* née au Royaume-Uni entre 1930 et 1940. Un terme γ qui dépend exclusivement de l'année de naissance $t - x$ est donc ajouté pour indiquer cette influence de l'année de naissance sur le niveau de mortalité. La formulation générale [voir Renshaw and Haberman (2006)] est la suivante :

$$\ln q_x(t) = \alpha_x + \beta_x \kappa_t + \gamma_{t-x} + \epsilon_{x,t}.$$

Un point essentiel dans ces modèles avec cohorte est l'identification de l'effet cohorte, puisque la date de naissance est égale à la différence entre la date courante et l'âge de l'individu [voir des discussions de ce problème d'identification dans Kuang et al. (2008), Mammen et al. (2011)].

D'un point de vue méthodologique, l'étude de la mortalité fait partie d'une plus large littérature sur les données de survie. Mais le modèle de Lee-Carter s'intéresse essentiellement à une durée univarée, i.e. la durée de vie de l'individu. Or les questions liées à l'étude des causes de mortalité, à l'étude jointe de la dépendance et de la durée de vie font intervenir plusieurs événements et nécessitent des modèles de survie bivarée. Un problème similaire existe dans beaucoup d'autres domaines : Par exemple :

— Dans un portefeuille de prêts hypothécaires , la fin du prêt peut être due à une défaillance de l'emprunteur, à un remboursement anticipé, à une re-négociation du contrat, ou à un

refinancement [voir Deng et al. (2000)].

— Quand la variable de durée est la durée de chômage d'un individu, la fin du chômage peut être due à un nouvel emploi, à une entrée dans une formation, à l'arrêt définitif de la recherche d'emploi, ou un passage à la retraite.

— Lorsque l'on étudie la survie des fonds spéculatifs, les informations disponibles sont auto-déclarées par les gestionnaires de fonds. Dans ce cas, la durée de vie du fonds est la différence entre la date de la dernière déclaration et la date d'émission. Cette fin de déclaration peut être due soit à une fermeture du fonds aux nouveaux investisseurs, soit à une liquidation suite à des performances trop décevantes [voir par exemple Haghani (2014)].

Dans la section suivante, nous introduisons la notion générale de modèle de survie bivariée.

I.3 Analyse de survie bivariée

Pour chaque individu, notons T_1 et T_2 les temps potentiels d'arrivée des deux événements. Donnons maintenant quelques exemples dans lesquels l'analyse des différents événements est importante et correspond à diverses situations d'observabilité.

Cas 1 : observations complètes. Dans ce cas, à la fois T_1 et T_2 sont observables. Par exemple,

— T_1 est le temps de décès de l'époux ;

— T_2 est le temps de décès de l'épouse.

Ici un "individu" est un couple et $T_1 > t, T_2 > t$ signifie que les deux époux sont vivants.

Case 2 : risques semi-concurrents. Dans ce cas , l'individu peut potentiellement rencontrer à la fois un événement non terminal (l'entrée en dépendance) et un événement terminal (la mort). Si l'événement terminal se produit en premier, l'événement non terminal n'est pas observé. Dans le cas contraire, nous observons les deux événements. Par exemple,

— T_1 est le temps potentiel d'entrée en dépendance ;

— T_2 est le temps de décès.

Dans cet exemple, $T_1 > t, T_2 > t$ signifie que l'individu est "vivant et autonome" et les deux risques sont dits semi-concurrents [voir e.g. Xu et al. (2010)].

Case 3 : risques concurrents. Les variables de durées latentes sont :

— T_1, le temps potentiel de décès dû à la cause 1 ;

— T_2, le temps potentiel de décès dû à la cause 2.

En pratique, on observe la date de décès :

$$T = \min(T_1, T_2),$$

ainsi que la cause du décès

$$J = 1 + \mathbb{1}_{T_1 > T_2}.$$

En d'autres termes, $J = 1$ (resp. $J = 2$) si et seulement si $T_1 < T_2$ (resp. $T_1 > T_2$).

Dans cet exemple, les variables T_1 and T_2 sont fondamentalement latentes car pour chaque individu, seulement l'une d'entre elles est observable.

La littérature de la survie introduit aussi des covariables observables, ainsi que des facteurs latents stochastiques pour prendre en compte la "corrélation" des événements. Dans la section suivante, nous discutons de façon plus précise les différents types de facteurs latents utilisés.

I.4 Facteurs latents

Il existe plusieurs types de facteurs latents avec des interprétations différentes.

i) **Facteurs individuels statiques** [voir e.g. Lancaster (1979)]. Ils sont également appelées hétérogénéité non observable, ou fragilité (frailty). Dans un modèle de survie univariée, l'introduction d'un facteur d'hétérogénéité a pour but de prendre en compte le biais de dépendance négative. Dans le cas bivarié, nous introduisons souvent un facteur d'hétérogénéité pour chaque variable de survie. Ces deux facteurs d'hétérogénéité peuvent être dépendants, ce qui rendra aussi les deux variables de survie dépendantes. Par exemple, il est généralement admis que le risque de décès dû au cancer est positivement corrélé avec le risque de décès dû aux maladies cardiovasculaires, car les deux types de maladies partagent des facteurs de risque communs (fumeur/non fumeur, niveau de pollution d'un pays, etc), dont tous ne sont pas observables.

ii) **Facteurs individuels dynamiques.** Le facteur individuel peut aussi dépendre du temps d'une manière stochastique. Par exemple, un individu peut entrer en dépendance avant le décès, et si cette entrée n'est pas observée par le statisticien, alors l'état dépendance/non dépendance

de cet individu est un facteur latent [voir le chapitre III pour une discussion plus détaillée de ce problème].

iii) **Facteurs communs dynamiques.** Traditionnellement dans les modèles de crédit, ils sont appelés fragilités dynamiques (dynamic frailty), [voir Duffie et al. (2009)]. Par exemple, dans le modèle de Lee-Carter, le facteur κ_t est un facteur latent dynamique ; on peut aussi imaginer que l'intensité d'entrée en dépendance et la mortalité avec ou sans dépendance diminuent toutes les trois, à cause d'un phénomène commun de longévité.

En pratique, il faut souvent introduire à la fois un facteur individuel (soit statique, soit dynamique) et un facteur commun dynamique. Dans ce cas là, la prise en compte simultanée de ces facteurs est cruciale pour ne pas avoir des résultats d'estimation erronés. Par exemple, on observe que durant les quarante dernières années, les taux de mortalité dus au cancer n'ont pas beaucoup diminué, malgré les progrès scientifiques dans la lutte contre le cancer. Comme expliqué dans Honoré and Lleras-Muney (2006), ce manque de diminution peut s'expliquer par la dépendance au niveau individuel entre les maladies cardiovasculaires et les cancers. Intuitivement, les individus ayant une plus grande intensité de décès due aux maladies cardiovasculaires ont aussi une plus grande probabilité d'avoir un cancer. Par conséquent, la forte diminution des taux de décès dus aux maladies cardiovasculaires ont eu un impact négatif sur le risque de mortalité du au cancer. Autrement dit, la diminution de ce dernier est partiellement “cachée" par cet effet négatif qu'il faut prendre en compte dans la modélisation.

La flexibilité des modèles à facteurs latents a une contrepartie : leur estimation est souvent difficile, notamment l'estimation de la distribution de l'hétérogénéité non observée. En effet, la théorie apporte peu d'information *a priori* sur la forme de cette distribution. Ainsi, il est souvent recommandé d'utiliser des estimateurs non-paramétriques de la distribution d'hétérogénéité, pour éviter l'introduction d'hypothèses paramétriques, qui peuvent être trop restrictives. Cette difficulté à estimer la distribution de l'hétérogénéité de manière précise a été documentée par Heckman and Singer (1984); Baker and Melino (2000) pour le cas univarié, et cette question est encore plus délicate dans les modèles de durée multivariés. Par conséquent, il y a souvent compromis entre flexibilité et robustesse. Le chapitre IV propose une nouvelle spécification de la loi du facteur individuel bivarié dans les modèles à facteur latent proportionnel.

Nous avons décrit, dans les deux dernières sections, les modèles traditionnels de survie bivariée. Néanmoins, même si les modèles de mortalité sont similaires aux modèles généraux de survie avec facteurs latents (individuels ou commun, dynamiques), la prise en compte de la longévité introduit de nouvelles problématiques. Par exemple,

— dans la littérature actuelle, les facteurs communs stochastiques ont été seulement introduits pour les modèles de survie univariés [voir e.g. Duffie et al. (2009)], sans tenir compte des facteurs d'hétérogénété individuelle. Or en assurance, il est souvent nécessaire d'introduire ces deux types de facteurs latents.

— il peut être utile de considérer, dans l'étude des événements sur plusieurs individus, des réactions asymétriques d'un individu au décès de l'autre. L'hypothèse de symétrie est habituellement faite en risque de crédit, lorsque les firmes considérées sont de même type de taille. Nous verrons une discussion plus détaillée de cette question dans le chapitre II "Love and Death : A Freund Model with Frailty".

— dans l'analyse de la longévité, le facteur commun est non stationnaire, alors qu'il est habituellement supposé stationnaire en risque de crédit. Ceci sera étudié dans le chapitre III "Long-Term Care and Longevity".

— en l'état actuel, certaines bases de données de mortalité ou de dépendance peuvent être difficilement utilisables, ce qui nous invite à proposer de nouvelles méthodologies indirectes d'analyse (voir par exemple le chapitre III "Long-Term Care and Longevity").

— en assurance-vie, les risques de long terme, par exemple la mortalité et l'intensité d'entrée en dépendance aux grands âges, ont une importance particulière pour les assureurs. Le chapitre IV "Large Duration Asymptotics" est dédié à cette discussion du risque aux grands âges en analyse de survie bivariée.

Les trois sections suivantes fournissent un résumé des trois chapitres suivants de la thèse.

I.5 Résumé du chapitre : "Love and Death : A Freund Model with Frailty"

Le second chapitre de cette thèse, intitulé "Love and Death : A Freund Model with Frailty", est basé sur un article du même nom, à paraître dans la revue *Insurance : Mathematics and Economics*.

Dans de nombreux pays, les produits joints d'assurance sont en train de gagner en popularité, surtout chez les couples retraités. Ces produits comprennent des assurances au dernier survivant (*last survivor*), des assurances décès écrites sur les deux têtes (*joint life*), ainsi que des rentes de réversion. Certains fonds de pension à bénéfice défini offrent également une clause de réversion, qui permet au conjoint survivant de l'employé de continuer à être couvert après le décès de cet employé.

Cet article étudie les liens entre les mortalités des deux époux d'un couple. Jusqu'à présent, la tarification des produits d'assurance ne tient pas compte de cette dépendance entre mortalités des époux. Nous montrons que ne pas tenir compte de cette dépendance peut entraîner des sur-évaluations ou sous-évaluations significatives au niveau des primes d'assurance.

La dépendance entre les mortalités des deux conjoints peut être de deux types. Tout d'abord, lorsque le premier conjoint décède, il peut y avoir une augmentation significative de la mortalité du conjoint survivant. C'est le syndrome du "coeur brisé" (*broken heart*). Par ailleurs, les deux conjoints partagent certains facteurs de risque communs, tels que le niveau d'éducation, la richesse, le style de vie, etc. Par conséquent, leurs états de santé sont positivement corrélés : une femme peu risquée est plus susceptible d'avoir un mari peu risqué.

Dans la littérature, la dépendance des mortalités a été préalablement modélisée soit par l'intermédiaire de copule, soit par des chaînes de Markov. Nous montrons que ces deux modèles ont des interprétations incompatibles. Plus précisément, le modèle de copule capture la dépendance due au facteur de risque commun, tandis que les chaînes de Markov capturent le syndrome du coeur brisé. Dans cet article, nous proposons un modèle qui permet de distinguer ces deux types de dépendance. Notre modèle englobe les deux modèles précédents comme cas particuliers. Nous expliquons également pourquoi il est possible d'identifier les paramètres des durées de vie des époux, sous des hypothèses raisonnables.

Enfin, pour illustrer les effets respectifs des deux types de dépendance sur la prime d'assurance, nous simulons une population hypothétique et calculons les taux de cotisation pour les différents produits d'assurance. Nous obtenons des structures de primes significativement différentes, si l'un de ces deux effets est ignoré. Nous étudions également la sensibilité des taux de prime aux valeurs des paramètres qui caractérisent ces deux types de dépendance. Nous concluons que l'absence de prise en compte de ces deux effets, ou la prise en compte à tort d'un seul d'entre eux, peut conduire à des sur-évaluations ou sous-évaluations importantes des primes d'assurance sur les produits écrits sur deux têtes.

I.6 Résumé du chapitre : “Long-Term Care and Longevity"

Ce chapitre introduit un nouveau modèle structurel de mortalité avec trois états : un état d'autonomie, un état intermédiaire, associé à une mortalité plus élevée (plus tard interprété comme l'état de dépendance), et un état de mort. Nous montrons que ce modèle est identifiable, tant que nous disposons des données par génération (cohorte) de mortalité et introduisons un facteur dynamique pour capturer le phénomène de longévité.

L'augmentation de la durée de vie espérée s'accompagne d'une augmentation du nombre de personnes âgées en perte d'autonomie, qui ont besoin de certaines formes de service de dépendance (long-term care, LTC). Une personne entre dans l'état LTC lorsqu'elle perd la capacité de marcher, de manger, de boire ou d'autres activités de la vie quotidienne. En raison du coût élevé de LTC, il est important d'analyser le temps passé dans cet état, ainsi que la probabilité d'entrer dans cet état, et comment ce temps et cette probabilité évoluent conjointement avec la longévité. Sont-ils presque indépendants de la longévité, ou augmentent-ils à un taux similaire ?

Cette analyse est souvent difficile à conduire en pratique à cause de la qualité des données de LTC. Tout d'abord, il n'y a pas une définition unique de l'état LTC, ce qui rend difficile la comparaison des différentes études existantes, en fait assez peu nombreuses. Deuxièmement, la collecte des données de dépendance est généralement difficile et imprécise. Troisièmement, même lorsque ces données existent, elles couvrent très peu de cohortes, ce qui empêche d'identifier les tendances sous-jacentes.

D'un autre côté, les données de mortalité sont beaucoup plus précises, cohorte par cohorte et facilement disponibles. Cet article développe un modèle qui nous permet de capturer l'évolution jointe des risque de dépendance et de mortalité, identifiable à partir des seules données de mortalité.

Plus précisément, nous caractérisons l'historique d'un individu en utilisant un modèle à trois états avec un état d'autonomie, un état latent de dépendance, et un état de mort. Dans ce modèle, la transition de l'état d'autonomie vers l'état de dépendance n'est pas observable, puisque nous n'observons que le décès. L'identification du modèle à partir de ces seules données de mortalité résulte de :

— l'hypothèse que l'entrée en dépendance entraîne une rupture dans le taux de mortalité.

— l'hypothèse d'un facteur commun dynamique de longévité, qui entraîne l'évolution de toutes les intensités de transitions entre différents états.

Ensuite, nous montrons comment estimer le modèle à partir des seules données de mortalité, et effectuons la prévision jointe des deux risques de mortalité et de dépendance.

I.7 Résumé du chapitre : "Large Duration Asymptotics in Bivariate Survival Models with Unobserved Heterogeneity"

Ce chapitre a deux contributions. Premièrement, il introduit un cadre pour étudier les propriétés aux grands âges (large duration asymptotics) des durées de vie résiduelles dans un modèle de survie à deux variables et à hétérogénéités proportionnelles. Ceci est important car le coût socio-économique associé à de très grandes valeurs de durée est élevé.

Bien que la théorie des valeurs extrêmes soit bien développée [voir e.g. Resnick (2007)], elle n'est pas directement utilisable pour les données de survie bivariée. En fait, les variables de durées bivariées possèdent beaucoup de caractéristiques spéciales :

— Elles sont souvent sujettes aux observations partielles (par exemple pour les risques concurrents). Par conséquent, certaines distributions marginales peuvent ne pas être observables. Dans ce cas là, les hypothèses sur les distributions marginales, standard dans la littérature des extrêmes, paraissent restrictives et inappropriées.

— Pour un même individu, différentes variables de durées partagent le même échelle de temps. Cela explique pourquoi les techniques de normalisation des marges utilisées en théorie des valeurs extrêmes sont en général peu appropriées en analyse de survie.

La deuxième contribution de ce chapitre est de donner les conditions assurant la convergence de la distribution des hétérogénéités parmi les survivants vers une distribution limite non dégénérée. Ceci généralise le résultat univarié établi par Abbring and van den Berg (2007). Cette famille de distributions limites est semi-paramétrique et donc plus parcimonieuse par rapport à une distribution non contrainte, qui est difficile à estimer. Elle est cependant suffisamment flexible par rapport à une distribution paramétrique. En effet, les distributions paramétriques actuellement utilisées sont souvent trop restrictives. Ceci est par exemple le cas de la distribution log-normale bivariée, proposée pour capturer la dépendance négative entre les différentes composantes d'hétérogénéité [voir Xue and Brookmeyer (1996)]. Par conséquent, elle est un concurrent sérieux des spécifications actuelles de l'hétérogénéité dans les modèles de survie bivariée.

Chapitre II

Love and Death : A Freund Model with Frailty

Abstract

We introduce new models for analyzing the mortality dependence between individuals in a couple. The mortality risk dependence is usually taken into account in the actuarial literature by introducing special copulas with continuous density. This practice implies symmetric effects on the remaining lifetime of the surviving spouse. The new model allows for both asymmetric reactions by means of a Freund model, and risk dependence by means of an unobservable common risk factor (or frailty). These models allow for distinguishing in the lifetime dependence the component due to common lifetime (frailty) from the jump in mortality intensity upon death of spouse (Freund model). The model is applied to the pricing of insurance products such as joint life policy, last survivor insurance, or contracts with reversionary annuities. A discussion of identification is also provided.

Keywords : Life Insurance, Coupled Lives, Frailty, Freund Model, Broken-Heart, Copula, Last Survivor Insurance, Competing Risks.

II.1 Introduction

This paper introduces new models for analyzing the mortality dependence between individuals in a couple. This type of model is needed for risk management and pricing of life insurance products written on two lives, such as joint life policy, last survivor insurance policy, or contract with reversionary annuities.

The basic actuarial literature usually assumed the independence between the spouses' mortality risks. Recently the mortality risk dependence has been introduced by means of copulas [see e.g. Frees et al. (1996), Youn and Shemyakin (1999), Carriere (2000), Denuit et al. (2001), Shemyakin and Youn (2006), Luciano et al. (2008), Luciano et al. (2010)], and the effect of this dependence on the risk premia starts to be measured. However, standard copula models assume continuous copula densities. This implies symmetric reactions of the mortality of a member of the couple when the other dies. An alternative consists in introducing jumps in mortality intensity (the Freund model) at the time of death of the spouse, to capture the death of a spouse [see e.g. Spreeuw and Wang (2008), Ji et al. (2011), Spreeuw and Owadally (2013)]. Our paper extends this literature by mixing the Freund model, which allows for asymmetric reactions of the mortality intensities at a death event, with unobservable common factor (or frailty), which underlies many usual Archimedean copulas [1].

The basic Freund model and its properties in terms of conditional intensities are presented in Section 2. This model allows for jump in the mortality intensity of a given spouse when the other spouse dies. The magnitude of this jump and its variation with respect to the age of the couple is the basis for constructing a convenient association measure, useful to analyze the broken-heart syndrome. The Freund model is extended in Section 3 to include common unobserved static frailty. In particular we discuss the properties of Freund models with latent intensities which are exponential affine functions of the frailty. These models are used in Section 4 to derive the prices of various contracts written on two lives. We consider these prices at the inception of the contract as well as during its lifetime. We emphasize the effect of the dependence between the mortality risks of the two spouses on these prices. Section 5 concludes. Proofs are gathered in appendices and a discussion on the identification issues is provided in Appendix A.4.

1. More precisely Archimedean copulas with completely monotone generators [see McNeil and Nešlehová (2009)]

II.2 The basic Freund model

This type of model has been introduced by Freund (1961) to construct bivariate survival models for dependent duration variables, while still featuring the lack of memory property. It has been noted by Tosch and Holmes (1980) that such models have an interpretation in terms of latent variables. We follow this interpretation. The model is written for a given couple, without specifying the index of the couple and possibly its observed characteristics such as the birth dates of the spouses, the difference between their ages [Youn and Shemyakin (1999)], or their age at the time of their marriage or common law relationship. In the application, such static couple characteristics will be introduced to capture the generation effects. The analysis is in continuous time and the lifetime variables are continuous variables.

II.2.1 The latent model

Let us consider a given couple with two spouses 1 and 2. The potential lifetimes of individuals 1 and 2, when both are alive, are denoted by X_1 and X_2, respectively. To get a unique time origin for the two members of the couple, these latent lifetimes are measured since the beginning of the common life. A first individual in the couple dies at date $\min(X_1, X_2)$. He/she is individual 1 (resp. individual 2), if $\min(X_1, X_2) = X_1$ [resp. $\min(X_1, X_2) = X_2$]. After this event, there can be a change in the potential residual lifetime distribution of the surviving individual. The potential residual lifetime of individual 1 (resp. individual 2) after the death of individual 2 (resp. individual 1) is denoted by X_3 (resp. X_4).

The joint distribution of the four latent variables is characterized by

i) the joint survival function of (X_1, X_2) :

$$S_{12}(x_1, x_2) = \mathbb{P}[X_1 > x_1, X_2 > x_2]; \tag{II-1}$$

ii) the survival function of X_3 given $X_2 = \min(X_1, X_2) = z$:

$$S_3(x_3; z) = \mathbb{P}[X_3 > x_3 | X_2 = \min(X_1, X_2) = z]. \tag{II-2}$$

iii) The survival function of X_4 given $X_1 = \min(X_1, X_2) = z$:

$$S_4(x_4; z) = \mathbb{P}[X_4 > x_4 | X_1 = \min(X_1, X_2) = z]. \tag{II-3}$$

These three joint and conditional survival functions, defined on $(0, \infty)$, characterize the latent model for the analysis of the mortality in the couple. In this model there exist at least three generation effects corresponding to the generations of each spouse, and to the generation of the couple, respectively.

II.2.2 Individual lifetimes

Link between the individual lifetimes and the latent variables

The lifetimes of individuals 1 and 2 (since the beginning of the common life) are denoted by Y_1 and Y_2. They can be expressed in terms of the latent variables as :

$$\begin{cases} Y_1 &= X_1 \mathbf{1}_{X_1<X_2} + (X_2 + X_3)\mathbf{1}_{X_2<X_1} = \min(X_1, X_2) + X_3 \mathbf{1}_{X_2<X_1,}, \\ Y_2 &= X_2 \mathbf{1}_{X_2<X_1} + (X_1 + X_4)\mathbf{1}_{X_1<X_2} = \min(X_1, X_2) + X_4 \mathbf{1}_{X_1<X_2}. \end{cases} \tag{II-4}$$

This system can be partially solved. First, the X_1, X_2 variables are related to variables (Y_1, Y_2) :

$\min(Y_1, Y_2) = \min(X_1, X_2)$, and $Y_1 > Y_2$, if and only if $X_1 > X_2$.

Then the variables X_3 and X_4 can be deduced in some regimes [2] since :

$$X_3 \mathbf{1}_{Y_2<Y_1} = Y_1 - \min(Y_1, Y_2) \text{ and } X_4 \mathbf{1}_{Y_1<Y_2} = Y_2 - \min(Y_1, Y_2).$$

As noted in Norberg (1989), the observed model can be interpreted in terms of a chain with four possible states [3], that are :

— state 1 : both spouses are alive,

— state 2 : husband dead, wife alive,

— state 3 : husband alive, wife dead,

— state 4 : both spouses are dead,

2. There are two regimes, corresponding respectively to the cases $Y_1 < Y_2$ and $Y_2 < Y_1$.

3. In their analysis Ji et al. (2011) consider also the possibility of a direct transition from state 1 to state 4 to account for catastrophic events (car accidents, plane crash) implying simultaneous deaths. They use a 5-day cut-off to account for a possible lag in reporting.

and transitions can only arise between states 1 and 2, 1 and 3, 2 and 4, and 3 and 4. Since the mortality intensity of a spouse can depend not only on the current state, but potentially on the time elapsed since the death of the other spouse, we get an example of a semi-Markov chain.

The joint density function and its decomposition

The joint probability density function (pdf) of (Y_1, Y_2) is easily derived from the distribution of the latent variables. We have (see Appendix A.1) :

$$\begin{aligned} f(y_1, y_2) &= \left[-\frac{\partial S_{12}}{\partial x_1}(y_1, y_1)\right]\left[-\frac{\partial S_4}{\partial x_4}(y_2 - y_1; y_1)\right], \text{ if } y_2 > y_1, \\ &= \left[-\frac{\partial S_{12}}{\partial x_2}(y_2, y_2)\right]\left[-\frac{\partial S_3}{\partial x_3}(y_1 - y_2; y_2)\right], \text{ if } y_1 > y_2. \end{aligned} \tag{II-5}$$

Therefore, the joint density function can feature a discontinuity when $y_1 = y_2$.

Let us consider the case $y_2 > y_1$. The density can also be written as :

$$f(y_1, y_2) = -\frac{\partial S^*}{\partial y}(y_1)\left[\frac{\partial S_{12}}{\partial x_1}(y_1, y_1)/\frac{\partial S^*}{\partial y}(y_1)\right]\left[-\frac{\partial S_4}{\partial x_4}(y_2 - y_1; y_1)\right], \tag{II-6}$$

where $S^*(y) = S_{12}(y, y)$ is the survival function of $\min(X_1, X_2)$ and
$\frac{\partial S^*}{\partial y}(y) = \frac{\partial S_{12}}{\partial x_1}(y, y) + \frac{\partial S_{12}}{\partial x_2}(y, y)$. Thus, the decomposition of the bivariate density involves three components :

i) $\left[-\frac{\partial S^*}{\partial y}(y_1)\right]$ is the density of the first death event ;

ii) the ratio $\left[\frac{\partial S_{12}}{\partial x_1}(y_1, y_1)/\frac{\partial S^*}{\partial y}(y_1)\right]$ is the probability that individual 1 dies at this first death event. It is equal to :

$$\mathbb{P}[Y_1 < Y_2 | \min(Y_1, Y_2) = y_1],$$

iii) $\left[-\frac{\partial S_4}{\partial x_4}(y_2 - y_1; y_1)\right]$ is the density of the residual lifetime after this event.

Individual mortality intensities

Let us now derive the individual mortality intensities given the current information concerning the couple. Their expressions depend on the state either alive, or dead, of the other spouse.

i) Let us first consider a date y at which both individuals are still alive, that is, such that $Y_1 \geq y, Y_2 \geq y$. The mortality intensity of individual 1 is defined by :

$$\begin{aligned} \lambda_1(y|Y_1 \geq y, Y_2 \geq y) &= \lim_{dy \to 0^+} \left\{ \frac{1}{dy} P[y \leq Y_1 \leq y + dy | Y_1 \geq y, Y_2 \geq y] \right\} \\ &= \int_y^\infty f(y, y_2) dy_2 / S^*(y). \end{aligned} \tag{II-7}$$

After replacing the bivariate density by its expression (2.5) for $y_2 > y_1$ and computing the integral, we get :

$$\lambda_1(y|Y_1 \geq y, Y_2 \geq y) = \left[-\frac{\partial S_{12}}{\partial x_1}(y, y) \right] / S^*(y). \tag{II-8}$$

This is the crude intensity function of individual 1 involved in the decomposition of the joint density function.

Similarly, we have :

$$\begin{aligned} \lambda_2(y|Y_1 \geq y, Y_2 \geq y) &= \lim_{dy \to 0^+} (\frac{1}{dy} P[y \leq Y_2 \leq y + dy | Y_1 \geq y, Y_2 \geq y]) \\ &= \int_y^\infty f(y_1, y) dy_1 / S^*(y). \\ &= \left[-\frac{\partial S_{12}}{\partial x_2}(y, y) \right] / S^*(y). \end{aligned} \tag{II-9}$$

ii) The expression of the mortality intensities can change if one of the individual dies exactly at date y. The mortality intensity of individual 1 at date y, if individual 2 dies at date y, becomes :

$$
\begin{aligned}
& \lambda_{1|2}(y|Y_1 \geq y, Y_2 = y) \\
= & \lim_{dy \to 0^+} \left[\frac{1}{dy} P(y < Y_1 \leq y + dy | Y_1 \geq y, Y_2 = y) \right] \\
= & \left[f(y,y) \right] / \left[-\frac{\partial S_{12}}{\partial x_2}(y,y) \right] \\
= & -\frac{\partial S_3}{\partial x_3}(0,y), \qquad \text{(II-10)}
\end{aligned}
$$

by applying the expression of the joint density (2.5) with $y_1 = y_2 = y$.

Similarly, we get :

$$
\begin{aligned}
& \lambda_{2|1}(y|Y_1 = y, Y_2 \geq y) \\
= & \lim_{dy \to 0^+} \left\{ \frac{1}{dy} P[y \leq Y_2 \leq y + dy | Y_1 = y, Y_2 \geq y] \right\} \\
= & -\frac{\partial S_4}{\partial x_4}(0,y). \qquad \text{(II-11)}
\end{aligned}
$$

Note that $S_3(0,y) = S_4(0,y) = 1$. Therefore we also have :

$$
\begin{aligned}
\lambda_{1|2}(y|Y_1 \geq y, Y_2 = y) &= -\frac{\partial \log S_3}{\partial x_3}(0,y), \\
\text{and } \lambda_{2|1}(y|Y_1 = y, Y_2 \geq y) &= \frac{-\partial \log S_4}{\partial x_4}(0,y),
\end{aligned}
$$

which are the expected expressions of the intensities in terms of survival functions.

iii) Finally, we can also consider the mortality intensity of spouse 1, when the other spouse is dead since a given time. We have, for $y > y^*$:

$$
\begin{aligned}
& \lambda_{1|2}(y|Y_1 \geq y, Y_2 = y^*) \\
= & \lim_{dy \to 0^+} \frac{1}{dy} P[y < Y_1 < y + dy | Y_1 \geq y, Y_2 = y^*] \\
= & f(y,y^*) / \int_y^\infty f(u,y^*) du \\
= & -\frac{\partial \log S_3}{\partial x_3}(y - y^*, y^*),
\end{aligned}
$$

which is just the intensity of the residual lifetime X_3 given the date of the first death.

Dependence and Jump in Intensities

It has been suggested in Clayton (1978) to measure the dependence between duration variables by considering the jump in intensities following the news of a death. We get a functional measure of dependence function of the age y of the couple, which is especially appropriate for following the dependence phenomenon during the couple life. These per-cent jumps are the following ones :

When individual 2 dies at date y, the jump at this date of the mortality intensity of individual 1 is :

$$\begin{aligned} \gamma_{1|2}(y) &= \lambda_{1|2}(y|Y_1 \geq y, Y_2 = y)/\lambda_1(y|Y_1 \geq y, Y_2 \geq y) \\ &= \left\{ \left[-\frac{\partial S_3}{\partial x_3}(0;y) \right] S^*(y) \right\} / \left[-\frac{\partial S_{12}}{\partial x_1}(y,y) \right]. \end{aligned} \tag{II-12}$$

Symmetrically, we get :

$$\begin{aligned} \gamma_{2|1}(y) &= \lambda_{2|1}(y|Y_1 = y, Y_2 \geq y)/\lambda_2(y|Y_1 \geq y, Y_2 \geq y) \\ &= \left\{ \left[-\frac{\partial S_4}{\partial x_4}(0;y) \right] S^*(y) \right\} / \left[-\frac{\partial S_{12}}{\partial x_2}(y,y) \right]. \end{aligned} \tag{II-13}$$

In the standard literature on bivariate survival models, the bivariate density function is continuous at $y_1 = y_2 = y$. Then, the two measures $\gamma_{1|2}(y)$ and $\gamma_{2|1}(y)$ coincide for any age y and it is easily checked that in this case, they are equal to the cross ratio function defined in Oakes (1989) [see also the discussion in Section 3.2]. This equality is not necessarily satisfied in a Freund model and we can observe different reactions of a spouse at the death of the other spouse in the couple.

Definition II.1. We have the immediate broken-heart syndrome for spouse 1 (resp. 2) at date y, if $\gamma_{1|2}(y) > 1$ [resp.$\gamma_{2|1}(y) > 1$].

We can have the immediate broken-heart syndrome (or the reverse immediate broken-heart syndrome when the directional measure of association is strictly smaller than 1), with different magnitude according to the age and spouse. We can even observe reactions in different directions. This arises when the wife is devastated by the death of her husband, with an increase of her mortality intensity, whereas the death of the wife may provide more freedom to her husband and

possibly a decrease of his mortality rate. This is the “love and death" phenomenon with the fact that love is not always shared and can be age-dependent.

Definition II.1 focuses on the immediate effect of the death of a spouse. According to this definition, many standard copula models [see e.g. Frees et al. (1996), Carriere (2000)] as well as the multiple state models in Ji et al. (2011) and Spreeuw and Owadally (2013) all allow for the broken-heart syndrome. There exist alternative definitions measuring the long-term or short-term persistence of the effect of the bereavement. For instance, Hougaard (2000) defines the broken-heart syndrome as a typical example of short-term effect : the mortality of the surviving spouse as a function of time elapsed since death of the partner is decreasing. Moreover, there can also be a long-term effect, that is, the effect of the death of the spouse is asymptotically non vanishing, or even increasing in the time elapsed. The Freund model, as well as models in Ji et al. (2011) and Spreeuw and Owadally (2013), are flexible enough to allow short-term (and/or long-term) effect ; on the other hand, Spreeuw (2006) shows that usual copula models can only capture long-term effect.

There exist a few studies trying to measure the effect and showing a positive estimated broken-heart syndrome [see e.g. Parkes et al. (1969), Jagger and Sutton (1991), Ji et al. (2011)]. Moreover it is shown that the broken-heart syndrome affects widowers more than widows [see Spreeuw and Owadally (2013)]. However, by neglecting the frailty effect discussed later on in Section 3, the estimates may suffer from an omitted heterogeneity bias.

II.2.3 Observed and latent intensities

Let us now link the distributions of the observed and latent variables. Since (X_1, X_3) and (X_2, X_4) cannot be simultaneously observed, let us first assume that these two pairs of variables are independent [4]. Then the distribution of the latent variables is characterized by the following latent intensities :

i) the latent intensity of X_1 denoted by $a_1(x_1)$;

ii) the latent intensity of X_2 denoted by $a_2(x_2)$;

iii) the latent intensity of X_3 given $X_2 = \min(X_1, X_2) = z$, denoted by $a_3(x_3; z)$;

iv) the latent intensity of X_4 given $X_1 = \min(X_1, X_2) = z$, denoted by $a_4(x_4; z)$.

4. In the next Section, this independence assumption is relaxed and replaced by an assumption of conditional independence given an unobserved heterogeneity variable F. Then by integrating out F, we will create unconditional dependence between the variables.

The associated cumulated intensities, that are their primitives with respect to the x argument, are denoted by $A_1(x_1), A_2(x_2), A_3(x_3; z), A_4(x_4; z)$, respectively. We deduce that :

$$S_{12}(x_1, x_2) = \exp\{-[A_1(x_1) + A_2(x_2)]\}, S_3(x_3; z) = \exp[-A_3(x_3; z)],$$
$$S_4(x_4; z) = \exp[-A_4(x_4; z)]$$

Then, the expression (2.5) of the bivariate probability density function becomes :

$$\begin{aligned} f(y_1, y_2) &= a_1(y_1) \exp\{-[A_1(y_1) + A_2(y_2)]\} a_4(y_2 - y_1; y_1) \exp[-A_4(y_2 - y_1; y_1)], \text{ if } y_2 > y_1, \\ &= a_2(y_2) \exp[-(A_1(y_1) + A_2(y_2))] a_3(y_1 - y_2; y_2) \exp[-A_3(y_1 - y_2; y_2)], \text{ if } y_1 > y_2. \end{aligned} \tag{II-14}$$

Similarly the directional measures of association can be written in terms of the latent intensities by using the expressions (2.12)-(2.13).

Property II.1. The directional measures of association are :

$$\gamma_{1|2}(y) = a_3(0; y)/a_1(y), \gamma_{2|1}(y) = a_4(0; y)/a_2(y). \tag{II-15}$$

II.3 Freund model with static frailty

The notion of (shared) frailty has been first introduced by Vaupel et al. (1979). The idea is to use the unobserved heterogeneity (or frailty) in bivariate duration models in order to create an additional dependence between lifetimes. In the basic specification, this frailty is static, since it depends on the couple only, neither on time, nor age. It represents the effect of common lifestyle, or common disasters encountered by the couple. In the extended model, the dependence between the lifetimes are due to either the frailty, or to the so-called contagion effects, that are the jumps in the intensities at the time of default. This new specification introduced below allows to disentangle these two effects. We first extend the Freund model of Section 2.4 to include unobserved frailty. Then, we discuss special cases.

II.3.1 The model

Let us denote by F the frailty variable, possibly multivariate. We consider a Freund model with the structure introduced in Section 2.4, where X_1 and X_2 are independent conditional on F, with latent intensities conditional on F given by : $a_1(x_1;F), a_2(x_2;F)$, $a_3(x_3;z;F), a_4(x_4;z,F)$. Let us now derive the latent [5] survival functions $S_{12}(x_1,x_2), S_3(x_3;z), S_4(x;z)$, when frailty F has been integrated out. We have :

$$\begin{aligned} S_{12}(x_1,x_2) &= \mathbb{E}\Big[\mathbb{P}[X_1 \geq x_1, X_2 \geq x_2|F]\Big] \\ &= \mathbb{E}\{\exp -[A_1(x_1;F) + A_2(x_2;F)]\}, \end{aligned}$$

where the expectation is taken with respect to the distribution of F.

Similarly we get :

$$\begin{aligned} S_3(x_3;z) &= \mathbb{P}[X_3 > x_3 | X_2 = \min(X_1,X_2) = z] \\ &= \mathbb{P}[X_3 > x_3 | X_2 = z, X_1 > z] \\ &= \frac{\mathbb{E}[a_2(z,F)\exp(-[A_1(z,F) + A_2(z;F) + A_3(x_3;z;F)])]}{\mathbb{E}[a_2(z;F)\exp(-[A_1(z;F) + A_2(z;F)])]}. \end{aligned}$$

These formulas can be used as inputs to derive the bivariate observed density (2.5) and the directional measures of association (2.12)-(2.13). For instance, we have by (2.12) :

$$\gamma_{1|2}(y) = \frac{\mathbb{E}\{a_3(0;y;F)a_2(y,F)\exp(-[A_1(y;F) + A_2(y;F)]\}\mathbb{E}[\exp(-[A_1(y;F) + A_2(y;F)])]}{\mathbb{E}\{a_2(y;F)\exp(-[A_1(y;F) + A_2(y;F)])\}E\{a_1(y;F)\exp[-A_1(y;F) + A_2(y;F)]\}}$$

We deduce the property below.

Property II.2.

$$\gamma_{1|2}(y) = \frac{\overset{Q_y}{\mathbb{E}}\ [a_3(0;y;F)a_2(y;F)]}{\overset{Q_y}{\mathbb{E}}\ [a_1(y;F)]\ \overset{Q_y}{\mathbb{E}}\ [a_2(y;F)]},$$

where Q_y denotes the probability distribution with density :

$$q_y(F) = \exp\{-[A_1(y) + A_2(y)]F\}/\mathbb{E}[\exp(-(A_1(y) + A_2(y))F],$$

with respect to the distribution of F. Thus, if the p.d.f. of F is $g(F)$, the p.d.f. of the modified measure Q_y is $q_y(F)g(F)$.

5. Note that the model has two layers of latent variables, first F, second X_1, X_2, X_3, X_4.

The change of density q_y is due to the aging of the heterogeneity structure in the population of surviving couples, called Population-at-Risk (PaR) at age y [see e.g. Vaupel et al. (1979), eq. (5)].

Since the conditional directional measure of association is [see (2.15)] :

$$\gamma_{1|2}(y;F) = a_3(0,y;F)/a_1(y,F),$$

we can also write the corresponding unconditional measure as :

$$\begin{aligned}
\gamma_{1|2}(y) &= \frac{\overset{Q_y}{\mathbb{E}}\left[\gamma_{1|2}(y;F)a_1(y;F)a_2(y;F)\right]}{\overset{Q_y}{\mathbb{E}}\left[a_1(y;F)\right]\overset{Q_y}{\mathbb{E}}\left(a_2(y;F)\right]} \\
&= \overset{\tilde{Q}_y}{\mathbb{E}}\left[\gamma_{1|2}(y;F)\right]\frac{\overset{Q_y}{\mathbb{E}}\left[a_1(y;F)a_2(y;F)\right]}{\overset{Q_y}{\mathbb{E}}\left[a_1(y;F)\right]\overset{Q_y}{\mathbb{E}}\left[a_2(y;F)\right]}, \\
\text{where : } d\tilde{Q}^y &= \frac{a_1(y;F)a_2(y;F)}{\overset{Q_y}{\mathbb{E}}\left[a_1(y;F)a_2(y;F)\right]}dQ^y.
\end{aligned}$$

Thus the unconditional directional measure of association $\gamma_{1|2}(y)$ is an average of the conditional directional measures of association with respect to a modified probability distribution, and adjusted for the dependence between $a_1(y;F)$ and $a_2(y;F)$, since the adjustment term equals 1, when these variables are not correlated under Q^y.

II.3.2 Single proportional frailty

Following Vaupel et al. (1979), it is usual to consider a single positive frailty with proportional effects on all latent intensities. This implies an Archimedean copula (with completely monotonic generator) for the bivariate latent variables X_1 and X_2 [see Oakes (1989), McNeil and Nešlehová (2009)], but not for the observed variables Y_1, Y_2, due to the changes in intensities after the first death event. More precisely, if :

$$a_1(x_1;F) = a_1(x_1)F, a_2(x_2;F) = a_2(x_2)F, a_3(x_3;z;F) = a_3(x_3;z)F; a_4(x_4;z;F) = a_4(x_4;z)F,$$

we deduce from Property II.2 that :

$$\gamma_{1|2}(y) = \frac{a_3(0;y)}{a_1(y)} \frac{\overset{Q_y}{\mathbb{E}}(F^2)}{[\overset{Q_y}{\mathbb{E}}(F)]^2}, \gamma_{2|1}(y) = \frac{a_4(0;y)}{a_2(y)} \frac{\overset{Q_y}{\mathbb{E}}(F^2)}{[\overset{Q_y}{\mathbb{E}}(F)]^2}.$$

In this simple case, the directional measures of association given F are [see (2.15)] :

$$\gamma_{1|2}(y;F) = \frac{a_3(0;y)F}{a_1(y)F} = \frac{a_3(0;y)}{a_1(y)}, \gamma_{2|1}(y;F) = \frac{a_4(0;y)}{a_2(y)}.$$

They are independent of frailty F, but not necessarily equal, which allows for asymmetric reactions.

The omitted heterogeneity introduces a positive bias on these measures. Indeed, we have $\overset{Q_y}{\mathbb{E}}(F^2)/[\overset{Q_y}{\mathbb{E}}(F)]^2 \geq 1$, by Cauchy-Schwartz inequality and more generally the property below :

Property II.3. In a Freund model with single proportional frailty the unconditional directional measures of association are larger than the conditional ones. They are equal if and only if frailty F is constant, that is, if there is no omitted heterogeneity :

$$\gamma_{1|2}(y) \geq \gamma_{1|2}(y;F), \gamma_{2|1}(y) \geq \gamma_{2|1}(y;F), \forall F.$$

However the per-cent adjustment for omitted heterogeneity is independent of age y and of the direction, which is considered. In particular the symmetry condition between spouses is preserved since :

$$\gamma_{1|2}(y;F) = \gamma_{2|1}(y;F) \Longleftrightarrow \gamma_{1|2}(y) = \gamma_{2|1}(y).$$

II.3.3 The actuarial literature

The models with mortality dependence considered in the actuarial literature are often special cases of the single proportional frailty model of Section 3.2.1, assuming moreover the continuity of the latent intensities :

Continuity assumption of the latent intensities

$$\begin{aligned} a_3(x_3;z) &= a_1(x_3+z), \forall x_3, z, \\ a_4(x_4;z) &= a_2(x_4+z), \forall x_4, z. \end{aligned}$$

Under the continuity assumption, the lifetimes Y_1, Y_2 are independent given the shared frailty F, with joint conditional survivor function :

$$S_{12}(y_1, y_2|F) = \exp[-[A_1(y_1) + A_2(y_2)]F].$$

To ensure the positivity of the intensity, the frailty F has to be positive. Let us denote by ψ its Laplace transform defined for positive arguments u by :

$$\psi(u) = \mathbb{E}[\exp(-uF)].$$

By integrating out the frailty, we deduce the joint survivor function :

$$S_{12}(y_1, y_2) = \psi[A_1(y_1) + A_2(y_2)].$$

A similar computation can be performed to derive the marginal survivor functions. We get :

$$S_1(y_1) = \psi[A_1(y_1)], S_2(y_2) = \psi[A_2(y_2)].$$

Since the Laplace transform of F is continuous and strictly increasing, it is invertible. We deduce the expression of S_{12} in terms of S_1, S_2 and ψ :

$$S_{12}(y_1, y_2) = \psi[\psi^{-1}[S_1(y_1)] + \psi^{-1}[S_2(y_2)]]$$

This is the standard definition of a copula [Sklar (1959)] :

$$S_{12}(y_1, y_2) = C[S_1(y_1), S_2(y_2)],$$

with a survivor Archimedean copula [Genest and MacKay (1986)] :

$$C(u_1, u_2) = \psi[\psi^{-1}(u_1) + \psi^{-1}(u_2)],$$

Property II.4. Let us consider a Freund model with single proportional frailty. Under the continuity assumption, the dependence between the lifetime variables Y_1, Y_2 is summarized by an Archimedean copula with the Laplace transform of the frailty as the generator.

Conversely, most usual Archimedean copulas admit a frailty interpretation [6]. The actuarial literature has considered this special case [see Tables 1 and 2, for examples in the actuarial literature, and Nelsen (1999) for a rather extensive list of copulas] [7] with different choices of the marginal distributions of the lifetimes and of the copulas.

TABLE II.1: Selected Marginal Distribution

Gompertz	Frees et al., (1996), Carriere (2000), Youn and Shemyakin (2001) Luciano et al. (2008, 2010)
Weibull	Frees et al. (1996), Youn and Shemyakin (1999, 2001), Shemyakin and Youn (2006)

TABLE II.2: Selected Copula

Frank	Frees et al., (1996), Carriere (2000), Youn and Shemyakin (2001) Spreeuw (2006), Luciano et al. (2008, 2010)
Gumbel-Hougaard	Youn and Shemyakin (1999, 2001), Shemyakin and Youn (2006) Spreeuw (2006) , Luciano et al. (2008, 2010)
Clayton	Carriere (2000), Luciano et al. (2008, 2010), Spreeuw (2006)
4.2.20 Nelsen copula	Spreeuw (2006), Luciano et al. (2008, 2010)

A more recent literature [see e.g. Denuit and Cornet (1999), Spreeuw (2006), Spreeuw and Wang (2008), Ji et al. (2011), Spreeuw and Owadally (2013)] focus on the broken-heart syndrome, but without introducing frailty in the specification of the intensities. This literature also identifies another downside of the common copula approach. Indeed, Spreeuw (2006) shows that for most

6. Indeed the Archimedean copulas that admit this representation are those whose generator is completely monotone, see McNeil and Nešlehová (2009) for a characterization of Archimedean copulas.

7. Some authors consider non Archimedean copulas, for instance normal copulas in Carriere (2000) or some multiple parameter families in Luciano et al. (2010). However, these copulas are still continuous and thus do not allow for asymmetric reactions. For this reason we have not listed these examples.

common Archimedean copulas, the mortality of the surviving spouse as a function of time elapsed since death of the partner is increasing, which is not underpinned by empirical evidences [see Spreeuw and Owadally (2013) as well as Section 2.2.4 for a relevant discussion].

II.3.4 Affine intensity model

A simple extension of the bivariate survival model discussed in Section 3.2 is obtained by introducing an intercept in the basic proportional frailty model [the so-called Generalized Shared Frailty model developed in Iachine (2004) in a special case]. The specification becomes :

$$\begin{aligned} a_1(x_1;F) &= a_1(x_1)F + b_1(x_1), a_2(x_2;F) = a_2(x_2)F + b_2(x_2), \\ a_3(x_3;z;F) &= a_3(x_3;z)F + b_3(x_3;z), a_4(x_4;z;F) = a_4(x_4;z)F + b_4(x_4;z). \end{aligned}$$

This extended version allows for conditional directional measures of association $\gamma_{1|2}(y;F)$ and $\gamma_{2|1}(y;F)$ depending on frailty F, and leads to non Archimedean copulas, when considering the joint distribution of latent lifetimes X_1 and X_2.

The affine specification is likely the most appropriate one for representing the effect of common lifestyle F and especially the memory features. After the death of a spouse, we expect that the effect of common lifestyle will diminish and asymptotically vanish. Thus, we expect that the latent intensity $a_3(x_3;z)$ [resp. $a_4(x_4;z)$] is a decreasing function of x_3 (resp. x_4) tending to zero at infinity. Then functions b_3 and b_4 provide the limiting mortality intensity a long time after the death of the other spouse. See also Section 2.2.4 for a detailed discussion on the long-term and short-term effect of losing his/her partner.

Finally, this affine intensity models assumes implicitly no remarriage or new common law relationship of the surviving spouse. This assumption is rather realistic for our purpose, since the insurance policies of interest are generally taken by rather old couples to profit of tax reductions, or to provide a rent to the surviving spouse.

II.4 Pricing contracts on two lives

We will now derive the pricing formulas for insurance contracts written on two lives such as joint life policies, last survivor policies and policies with reversionary annuities. By considering extended Freund models (under the risk-neutral probability), we analyze the effect of jumps in intensity on prices at the contract issuing as well as on the premium updating during the life of

the contract.

II.4.1 Prices at the inception of the contracts

The premium computations for the joint policies are based on the joint remaining lifetimes risk-neutral distribution conditional on the ages of the spouses at the beginning of their common life $y^*_{1,0}, y^*_{2,0}$, say, and on the fact that both spouses are still alive with an age of the life in couple equal to z_0, say, at the inception of the contract. Thus, the joint risk-neutral density of the remaining lifetimes $\tilde{y}_j = Y_j - z_0, j = 1, 2$ at the inception of the contract is [8] :

$$\begin{aligned} & \tilde{f}_0(\tilde{y}_1, \tilde{y}_2 | z_0) \\ = & \lim_{dy_1, dy_2 \to 0} \left\{ \frac{1}{dy_1 dy_2} P[Y_1 \in (\tilde{y}_1 + z_0, \tilde{y}_1 + z_0 + dy_1), Y_2 \in (\tilde{y}_2 + z_0, \tilde{y}_2 + z_0 + dy_2) \right. \\ & \left. | Y_1 > z_0, Y_2 \geq z_0, y^*_{1,0}, y^*_{2,0}] \right. \\ = & f_0(\tilde{y}_1 + z_0, \tilde{y}_2 + z_0)/S_0(z_0), \end{aligned}$$

where the index 0 means that the distribution characteristics of Section 3 can now depend on the initial ages $y^*_{1,0}, y^*_{2,0}$.

Let us now illustrate the premium computation in a continuous time framework with instantaneous constant interest rate r. For each insurance product, we have to analyze the risk-neutral distribution of the discounted cash-flows.

i) Joint life policy

Let us denote by a the premium rate and consider an insurance paying 1$ immediately at the first death of a spouse. The discounted sequence of cash-flows measured at the inception of the contract is :

8. The link between the historical and risk-neutral bivariate distributions of the lifetimes is discussed in Appendix A.2. The insurance literature often prices the insurance contracts by means of the historical distribution to get the so called fair premium, that is, neglects the correction for risk [see e.g. Ji et al. (2011), Section 5.6].

$$
\begin{aligned}
C_0^{(1)}(a,r,z_0;Y_1,Y_2) &= a\int_0^{\min(Y_1,Y_2)-z_0} \exp(-rh)dh - \exp[-r(\min(Y_1,Y_2)-z_0)] \\
&= \frac{a}{r}\{1-\exp[-r(\min(Y_1,Y_2)-z_0)]\} - \exp[-r(\min(Y_1,Y_2)-z_0)]\}.
\end{aligned}
$$

There exist different ways for balancing the stochastic positive and negative cash-flows. In particular the premium rate [9] can be defined by fixing equal expectations to these sequences. We get :

$$
a_0^{*(1)}(r) = r\frac{\mathbb{E}_0\{\exp[-r(\min(Y_1,Y_2)-z_0)]|Y_1\geq z_0, Y_2\geq z_0\}}{1-\mathbb{E}_0\{\exp[-r(\min(Y_1,Y_2)-z_0)]|Y_1\geq z_0, Y_2\geq z_0\}}.
$$

ii) Last survivor policy

Let us now assume that the death event written in the policy is the second death of a spouse. The formulas are the same as for the joint life policy above after substituting $\max(Y_1,Y_2)$ to $\min(Y_1,Y_2)$. For instance, the fair premium becomes :

$$
a_0^{*(2)}(r) = r\frac{\mathbb{E}_0(\exp[-r(\max(Y_1,Y_2)-z_0)]|Y_1\geq z_0, Y_2\geq z_0)}{1-\mathbb{E}_0\{\exp[-r(\max(Y_1,Y_2)-z_0)]|Y_1\geq z_0, Y_2\geq z_0\}}.
$$

iii) Reversionary annuities

Finally, let us consider a product in which the premium is paid when both spouses are alive and a unitary annuity is paid to the surviving spouse up to his/her death. The discounted sequence of cash-flows becomes :

$$
\begin{aligned}
C^{(3)}(a,r,z_0;Y_1,Y_2) &= a\int_0^{\min(Y_1,Y_2)-z_0} \exp(-rh)dh - \int_{\min(Y_1,Y_2)-z_0}^{\max(Y_1,Y_2)-z_0} \exp(-rh)dh \\
&= \frac{a}{r}\{1-\exp(-r[\min(Y_1,Y_2)-z_0])\} \\
&\quad -\frac{1}{r}\{\exp[-r(\min(Y_1,Y_2)-z_0)] \\
&\quad -\exp[-r(\max(Y_1,Y_2)-z_0)]\}.
\end{aligned}
$$

9. The fair premium rate is obtained by replacing the risk-neutral distribution by the historical distribution in formula (4.3). Otherwise the premium rate accounts for a risk premium.

The associated premium rate is :

$$a_0^{*(3)}(r) = \frac{\mathbb{E}_0\{\exp(-r[\min(Y_1,Y_2)-z_0]) - \exp(-r[\max(Y_1,Y_2)-z_0])|Y_1 \geq z_0, Y_2 \geq z_0\}}{1-\mathbb{E}_0\{\exp(-r[\min(Y_1,Y_2)-z_0])|Y_1 \geq z_0, Y_2 \geq z_0\}}.$$

iv) Individual products

The premia for joint products have naturally to be compared with the premia of a life insurance paying 1$ at the death of a single life.

The associated fair premium is :

$$a_{j,0}^{*}(r) = r\frac{\mathbb{E}_0(\exp[-r(Y_j-z_0)]|Y_j \geq z_0)}{1-\mathbb{E}_0(\exp[-r(Y_j-z_0)]|Y_j \geq z_0])},$$

if only information on spouse j is taken into account and

$$a_{j,0}^{**}(r) = \frac{r\mathbb{E}_0(\exp[-r(Y_j-z_0)]|Y_1 \geq z_0, Y_2 \geq z_0)}{1-\mathbb{E}_0(\exp[-r(Y_j-z_0)]|Y_1 \geq z_0, Y_2 \geq z_0)},$$

if the information on the couple is taken into account.

In the limiting case of a zero risk-free rate $r = 0$, the expressions of the premia are obtained by a Taylor expansion. We get :

$$a_0^{*(1)}(0) = \frac{1}{\mathbb{E}_0\{[\min(Y_1,Y_2)-z_0]|Y_1 \geq z_0, Y_2 \geq z_0\}},$$

$$a_0^{*(2)}(0) = \frac{1}{\mathbb{E}_0\{[\max(Y_1,Y_2)-z_0]|Y_1 \geq z_0, Y_2 \geq z_0\}},$$

$$a_0^{*(3)}(0) = \frac{\mathbb{E}_0\{\max(Y_1,Y_2)-\min(Y_1,Y_2)|Y_1 \geq z_0, Y_2 \geq z_0\}}{\mathbb{E}_0\{\min(Y_1,Y_2)|Y_1 \geq z_0, Y_2 \geq z_0\}},$$

$$a_{j,0}^{*}(0) = \frac{1}{\mathbb{E}_0\{Y_j-z_0|Y_j \geq z_0\}},$$

$$a_{j,0}^{**}(0) = \frac{1}{\mathbb{E}_0\{Y_j-z_0|Y_1 \geq z_0, Y_2 \geq z_0\}}.$$

The pricing of the individual contracts of two spouses cannot be done separately. Indeed the

survival probabilities of a single life, and then the price of the individual contract, depend on the life history of the spouse, whether or not he/she is still alive and, when he/she died if applicable [see e.g. Youn et al. (2002)].

II.4.2 Effect of risk dependence on prices

Let us now illustrate the effect on policy prices of risk dependencies : due to the frailty and to the asymmetric jump in intensities existing in a Freund model.

We consider a model with single proportional frailty (see Section 3.2). The population of couples is such that the two spouses have the same age 30. The distribution of the heterogeneity F at age 30 is assumed to be a gamma distribution. Note that when there is no jump in latent intensities, the joint distribution of the lifetimes is associated to a Clayton copula. Due to the mover-stayer phenomenon, as the population ages, the distribution given that both spouses survive up to age $z_0 > 30$, that is, the heterogeneity distribution that the insurance company applies to price a contract for a couple with an underwriting age $z_0 > 30$, will depend on age z_0. Intensities of the latent duration variables X_1 (female), X_2 (male) are of the following form :

$$a_1(x_1) = \exp(\alpha_1 x_1 + \beta_1), \qquad \forall x_1 > 0,$$

and

$$a_2(x_2) = \exp(\alpha_2 x_2 + \beta_2), \qquad \forall x_2 > 0.$$

For illustration purpose, we assume that the death of the spouse has a constant multiplicative effect γ on the mortality intensity of the survivor. Thus, given $z = \min(X_1, X_2)$, the conditional intensities of X_3, X_4 are of the form :

$$a_3(x_3, z) = \gamma \exp\Big(\alpha_1(z + x_3) + \beta_1\Big), \qquad \forall x_3 > 0,$$

and

$$a_4(x_4, z) = \gamma \exp\Big(\alpha_2(z + x_4) + \beta_2\Big), \qquad \forall x_4 > 0,$$

where the constant $\gamma = \frac{a_3(0,z)}{a_1(z)} = \frac{a_4(0,z)}{a_2(z)}$ is larger than 1 to reflect the broken-heart syndrome. Thus the model adopted here is similar to Denuit and Cornet (1999) except that frailty is incorporated. For the illustration the jump in mortality on death of the first life is the same,

whether male or female. For numerical illustrations, parameters $\alpha_1, \alpha_2, \beta_1, \beta_2$ are chosen to fit the marginal intensities of American females and males at ages $31, 32, ..., 110$, provided by the Human Mortality Database [10]. Their values are reported below :

$$\alpha_1 = 0.089, \beta_1 = -7.613, \alpha_2 = 0.081, \beta_2 = -6.934.$$

The measure of association γ is the same in both directions with values $\gamma \in \{1, 3, 5\}$. $\gamma = 5$ corresponds to a very huge impact of the death of the spouse on the survivor lifetime and $\gamma = 1$ corresponds to the case of no impact (at the individual level, indeed, even in this case there is still jump of intensity when the heterogeneity is integrated out, see e.q.(3.2)). The gamma distribution of the heterogeneity at age 30 is set to have a shape parameter k and a scale parameter $1/k$. Therefore, the average mortality intensity at age 30 is the same for each value of k, since $\mathbb{E}(F) = 1/k \cdot k = 1$ does not depend on k. The heterogeneity parameter k will be set to $k \in \{2, 5, 10\}$. $k = 10$ corresponds to a low heterogeneity level and $k = 2$ corresponds to a high one. This specification of the duration distribution is the risk-neutral distribution, which can be used to price the different life insurance contracts described in Section 4.1. The risk-free interest rate is set to $r = 1\%$. We provide in Figure II-1 the evolution of the premium rates as a function of the underwriting age $z_0 \in 31, 32, ..., 80$, for different contracts and for $\gamma = 5, k = 2$. The contracts include a joint life policy, a last survivor policy, a contract with reversionary annuities, and the individual insurance products for female with, or without, the information on the survival of the husband up to z_0.

10. The Human Mortality Database (HMD) was created to provide detailed mortality and population data to researchers, students, journalists, policy analysts, and others interested in the history of human longevity. It is maintained by the University of California, Berkeley, and the Max Planck Institute for Demographic Research in Rostock, Germany ; its official website is http ://www.mortality.org

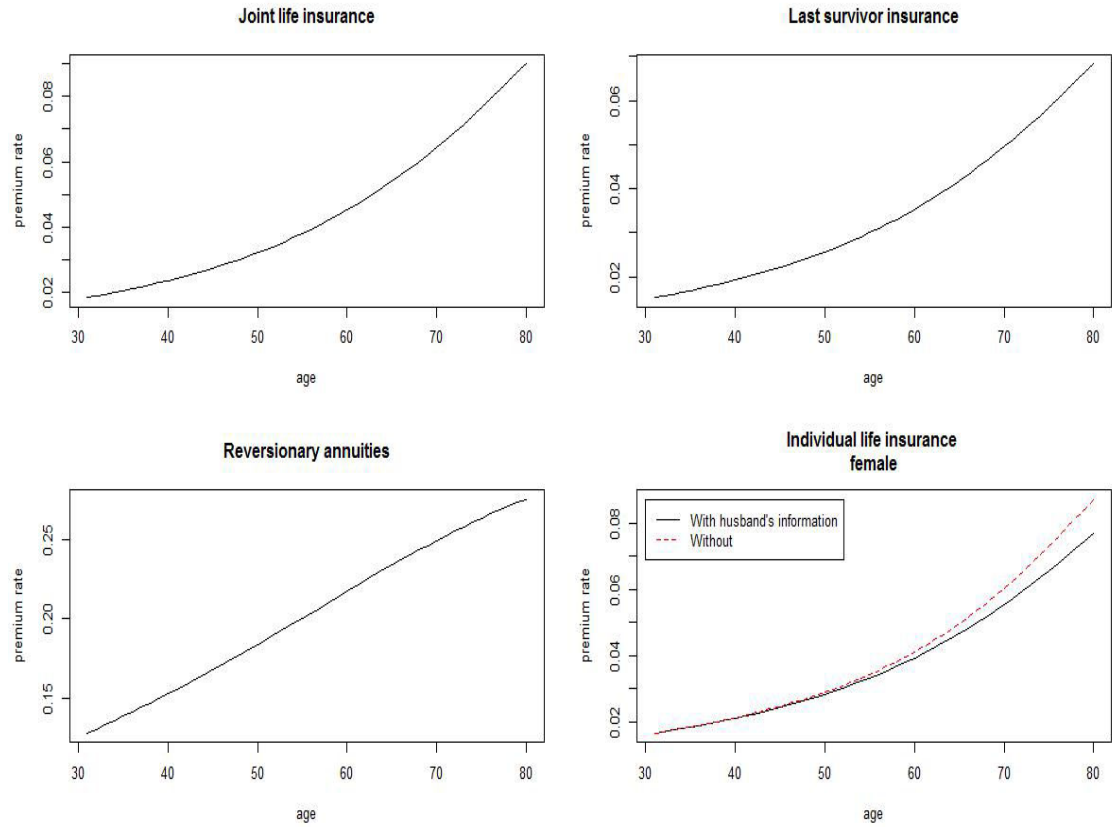


FIGURE II-1: Premium rate as a function of the age of the couple at the time of underwriting. In the lower right panel for individual life insurance policies, the dashed line (respectively solid line) represents the premium rates when the information on the spouse is (respectively is not) taken into account.

These premia are not directly comparable, since the premia paid by the insured people (resp. the payments by the insurance company) do not correspond to a same period. Nevertheless for each product, the premium rate is increasing with the age of underwriting of the couple, which is in conformity with the usual premium structure without heterogeneity.

In general, in a model with heterogeneity, the average intensity (as well as the premium) is not necessarily monotone in z_0. Indeed, the aging of the population has a positive impact on the premium when z_0 increases, while the mover-stayer phenomenon has a negative impact on the premium since couples with higher risks die out more quickly ; hence the average heterogeneity is improving in time. In this example, the first effect is more important, which results in an increasing premium.

Besides, the premium rate of an individual insurance contract for a female is always lower

when the insurance company know that her spouse is still alive, as shown in the lower right panel. The difference is negligible at low ages, but increases significantly with respect to z_0. We also observe that the curves of the premia are convex, except for reversionary annuities, where the trend is almost linear.

Let us now illustrate the effect of risk dependencies and heterogeneity for the different insurance contracts. We first illustrate in Tables II.3 and II.4 the effect of the measure of association γ for two different ages 30 and 50. This parameter has no effect on the joint insurance policies : indeed, the contract terminates up to the first death whereas the measure of association impacts only the residual lifetime beyond the first death event. Therefore, premium rates of the joint insurance are not reported in the Tables. The two last columns correspond to the individual insurance contract for a female with and without information on the survival of her spouse. We get premia, which increase with the γ parameter, except for the reversionary annuities. Indeed, unlike other contracts which concern death benefit, a reversionary annuity pays survival benefits ; therefore its relationship with the deterioration of mortality is opposite to other products.

	Last survivor	Reversionary annuity	Individual, female, without husband's information	Individual, female, with husband's information
$\gamma = 5$	0.0194	0.134	0.0212	0.0210
$\gamma = 3$	0.0182	0.181	0.0203	0.0202
$\gamma = 1$	0.0153	0.318	0.0184	0.0183

TABLE II.3: Effect of the broken heart syndrome on premium rates with a fixed heterogeneity distribution ($k = 6$), at age 30.

	Last survivor	Reversionary annuity	Individual, female, with husband's information	Individual, female, without husband's information
$\gamma = 5$	0.0279	0.166	0.0319	0.0303
$\gamma = 3$	0.0260	0.225	0.0309	0.0290
$\gamma = 1$	0.0214	0.404	0.0275	0.0258

TABLE II.4: Effect of the broken heart syndrome on premium rates with a fixed heterogeneity distribution ($k = 6$), at age 50.

Then we illustrate in Tables II.5 and II.6 the effect of heterogeneity, characterized by parameter k, for two different ages 30 and 50. For instance, for the joint life contract, the premium increases as the heterogeneity decreases [11]. However, this effect is less clear for other products.

11. This is expected. Indeed, the unconditional survivor function of the first death is :

$$S^*(t) = \mathbb{E}[e^{-(A_1(t)+A_2(t))F}] = \frac{1}{\Big(1 + 1/k(A_1(t) + A_2(t))\Big)^k},$$

Indeed, in a more heterogeneous population ($k = 2$), there are more couples of extremely high risk, as well as more couples of extremely low risk. The first couples contribute to an increase in the premium whereas the latter couples contribute to diminish the premium. For the reversionary annuity, a riskier couple is expected to trigger annuity payment earlier, which means less premium income, but the payment is also expected to terminate earlier, which spells less total payment. In our simulation studies, we observe that, for each product, the premium rate is decreasing in the heterogeneity, both for age 30 and 50. Figure II-2 plots, for each k, simulated lifetimes distributions for the last survivor, respectively for $z_0 = 30$ and 50.

	Joint life	Last survivor	Reversionary annuity	Individual, female, with husband's information	Individual, female, without husband's information
$k = 2$	0.0186	0.0153	0.129	0.0167	0.0167
$k = 6$	0.0196	0.0161	0.135	0.0176	0.0176
$k = 10$	0.0197	0.0162	0.136	0.0177	0.0177

TABLE II.5: Effect of heterogeneity on premium rates with a fixed broken heart syndrome ($\gamma = 5$), at age 30.

	Joint life	Last survivor	Reversionary annuity	Individual, female, with husband's information	Individual, female, without husband's information
$k = 2$	0.0334	0.0265	0.188	0.0299	0.0293
$k = 6$	0.0364	0.0287	0.199	0.0324	0.0318
$k = 10$	0.0371	0.0292	0.203	0.0329	0.0323

TABLE II.6: Effect of heterogeneity on premium rates with a fixed broken heart syndrome ($\gamma = 5$), at age 50.

and the corresponding unconditional intensity function is :

$$\lambda(t) = \frac{a_1(t) + a_2(t)}{1 + 1/k(A_1(t) + A_2(t))},$$

thus the premia for a joint life contract is higher for $k = 10$.

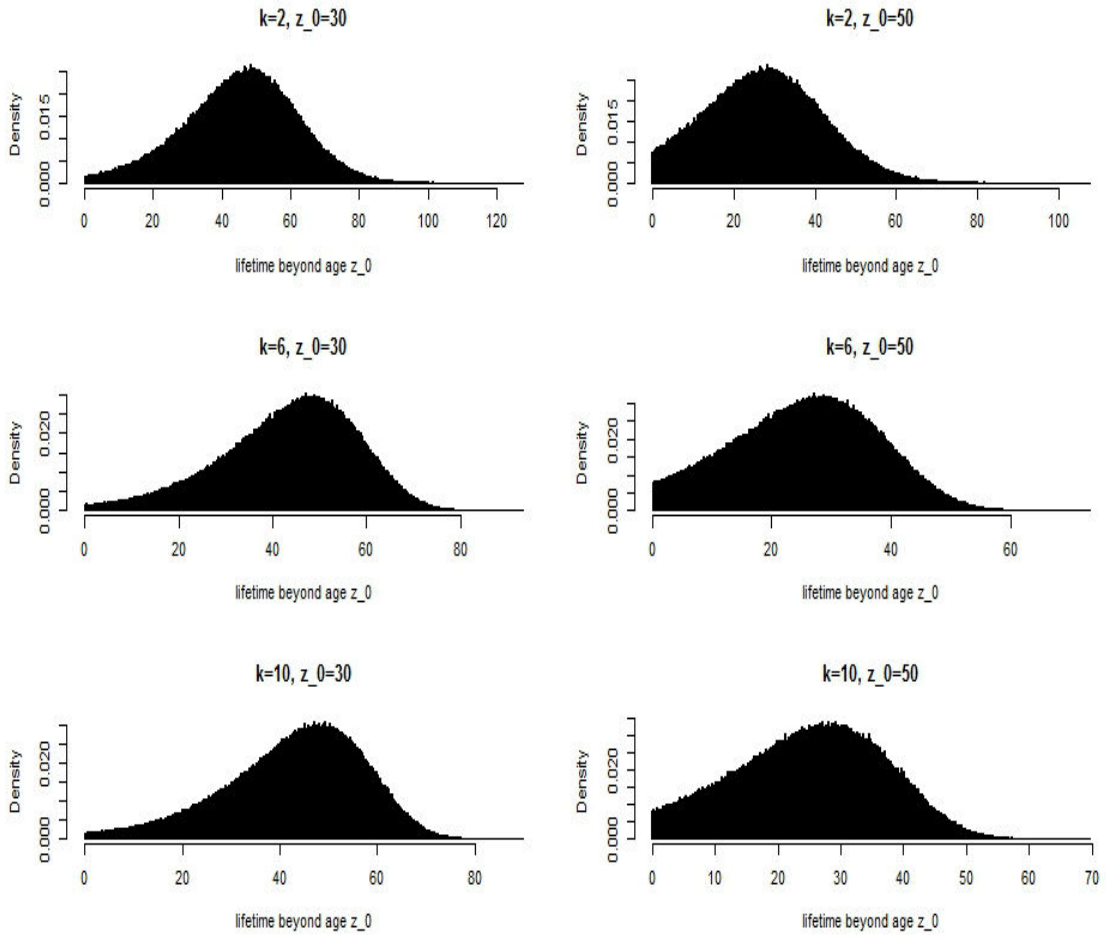


FIGURE II-2: Probability density functions of the last survivor's lifetime upon z_0, for $z_0 = 30, 50$.

Special attention should be paid when comparing premium rates at age 50 for different values of parameter k. Indeed, for each value of k, $\gamma(k, 1/k)$ is the heterogeneity distribution at age 30, but the heterogeneity distribution conditional on the survival of both spouses up to age 50 is no longer the same. However, it is still a gamma distribution $\gamma(k, 1/[k + A_1(z_1 - z_0) + A_2(z_1 - z_0)])$, where $z_0 = 30$, $z_1 = 50$ and A_1, A_2 are the cumulative intensities (see **??**). Therefore, the mean of the heterogeneity is $k/[k + A_1(z_1 - z_0) + A_2(z_1 - z_0)]$, and quotient between the variance at age 50 and that at age 30 is $k^2/[k + A_1(z_1 - z_0) + A_2(z_1 - z_0)]^2$. Both quantities are decreasing functions of k, that is, the mean and the variance of the heterogeneity diminish (in proportion) faster in the population with initially the highest heterogeneity ($k = 2$). Figure II-3 plots, for each k, the probability density function of the heterogeneity both at age 30 and at age 50. The gamma distribution parameters at age 50 are reported in Table II.7.

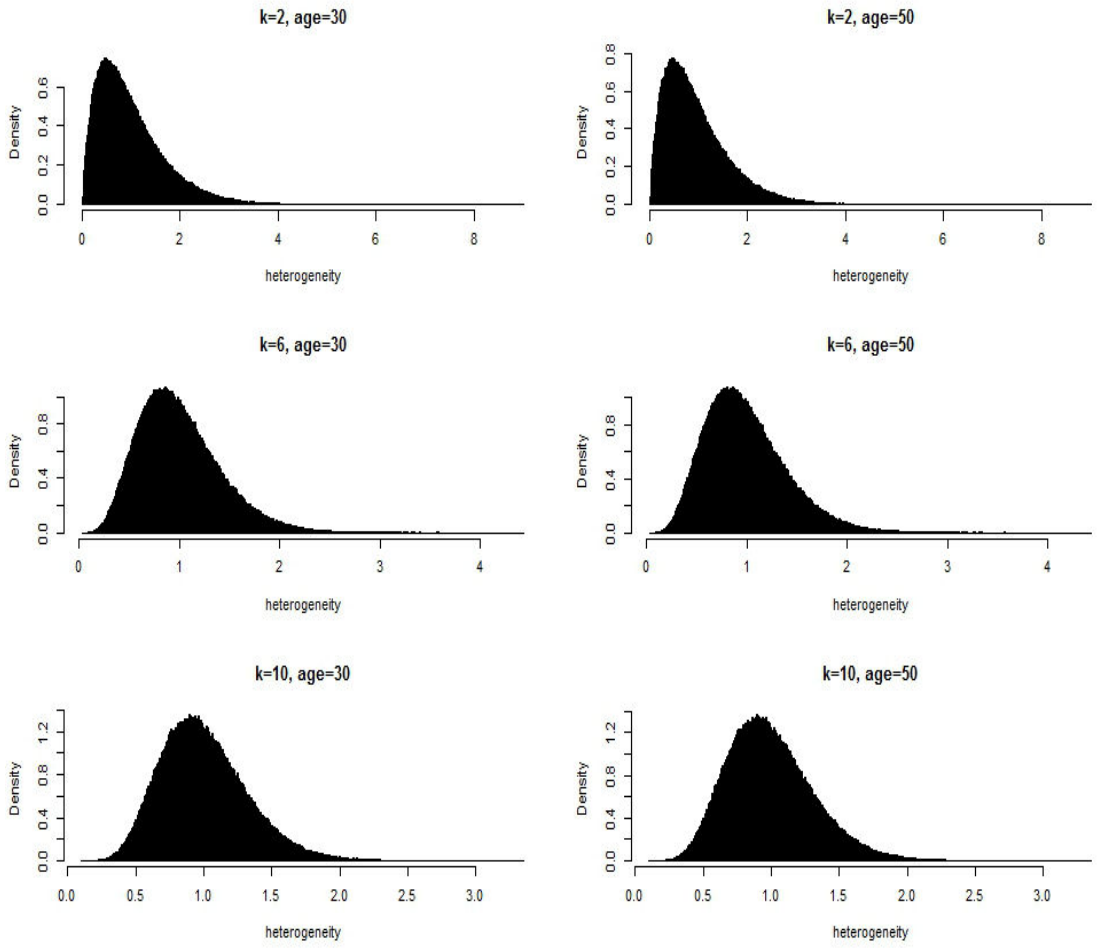


FIGURE II-3: Probability density functions of the heterogeneity, at ages 30 and 50.

	Shape parameter	Scale parameter	$\sqrt{\frac{\text{Variance at age 50}}{\text{Variance at age 30}}}$
$k = 2$	0.4816	2	0.9279
$k = 6$	0.1646	6	0.9750
$k = 10$	0.0992	10	0.9849

TABLE II.7: Gamma distribution parameters at age 50 for different gamma distributions $\gamma(k, 1/k)$ at age 30. The scale parameter is the same as at age 30. The fourth column gives values of $k/[k + A_1(x) + A_2(x)]$, which equals also the mean of the heterogeneity distribution. It measures the reduction of the heterogeneity due to the mover-stayer phenomenon.

II.4.3 Evolution of the price of the contract during the life of the contract

A premium level a_0 is fixed at the inception of each contract (see Section 4.1). However, it is important to evaluate regularly the residual value of this contract during its life, for instance, to include it correctly in the balance sheet, or, if it is securitized, to evaluate the price of the corresponding component of the Insurance Linked Security (ILS).

Let us first focus on the joint life policy. The fair value of this contract at a date where both spouses are still alive and the age of the couple is $z_1, z_1 \geq z_0$, is given by :

$$\begin{aligned} & C^{(1)}_{1|0}(a_0, r, z_1; Y_1, Y_2) \\ = \quad & \mathbb{E}_0[C^{(1)}_0(a_0, r, z_1; Y_1, Y_2)|Y_1 \geq z_1, Y_2 \geq z_1]. \end{aligned}$$

a_0 is for instance equal to the fair premium $a_0 = a_0^{*(1)}$ given in (4.3) when $z_1 = z_0$.

The price updating is more complicated for the reversionary annuities product, since we have to distinguish the two possible regimes existing during the life of the contract. In the first regime the two spouses are both alive, with an age of the couple equal to z_1. In the second regime, there is just one surviving spouse, the available information includes the date of the first death and the fact that the surviving spouse is the husband, or the wife. In both regimes, the residual value is systematically negative. First, in the second regime the only cash flows are the payment of the annuity, which are negative. Second, in the first regime, the premium rate of the reversionary annuity is increasing in z_0 (see Figure II-1), therefore, couples who entered into the contract at age $z_0 < z_1$ pay, at age z_1, less premium than newly underwritten couples of age z_1, while the two groups have the same heterogeneity distribution, thus the same risk profile.

For illustration, let us calculate the residual value of a reversionary annuity underwritten at the age of 30. At date $t > 30$, the residual value of this contract depends on the survival status of the couple. We use the same model as in the previous section and Figure II-4 displays the evolution of the residual value of the contract, first when both spouses are still alive at date t, then when one of the spouse died before t. The parameters are $\gamma = 5, k = 2, z_0 = 30$. As expected we observe that in both case, the value of the contract is negative. We observe also in the second case, that the value of the contract is smaller for widows than for widowers. Indeed, at the same

age and with the same marital status, women have a smaller mortality intensity than men have.

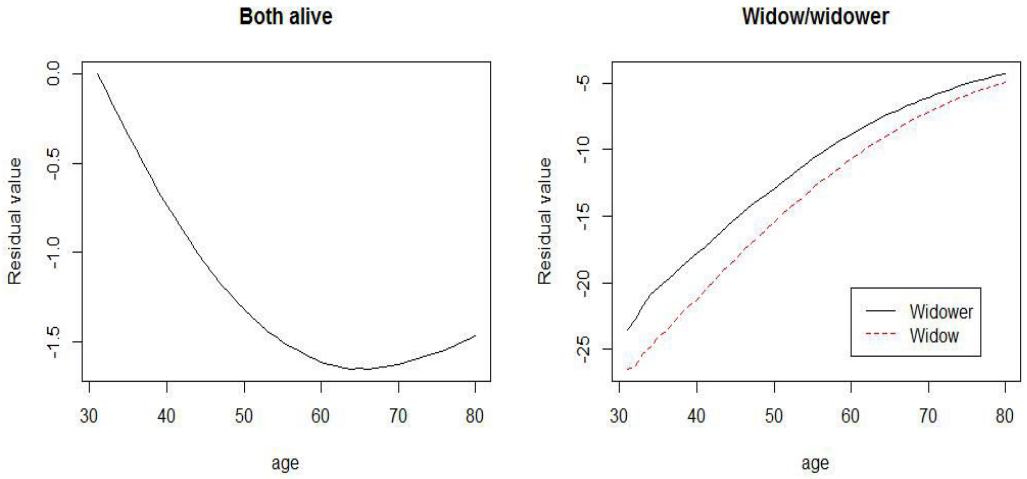


FIGURE II-4: Evolution of the residual value of a reversionary annuity. Left panel : both spouses are still alive. Right panel : one of the spouses died before t.

II.5 Concluding remarks

The standard insurance literature for analyzing and pricing insurance contracts written on two lives are pure models. A first category assumes a continuous bivariate distribution of the spouses' lifetimes with a continuous probability density function. This continuity assumption implies no jump in intensity when a spouse dies. A second category of models apply a pure Freund model to describe the broken-heart syndrome. These two effects impact the price of insurance contracts and of annuity values in different ways, not only the price of contracts written on two lives, but also the prices of individual contracts written on a single life [12]. By considering appropriate extensions of the Freund model, we have explained how to account for both individual heterogeneity and potential jumps at the time of a spouse's death.

A similar problem arises in the credit risk literature where the death event is replaced by a default event. The standard credit risk literature prices the default intensity, not the default event itself, leading to possible mis-pricing of credit derivatives. The idea of introducing jumps in intensity to correct such a mispricing has been proposed in Jarrow and Yu (2001) for a credit

12. For the same reason they can impact the price of health insurance or of long-term care contracts, for instance, since the risk of entering into long-term institutional care after the death of a spouse can increase [Nihtilä and Martikainen (2008)].

derivative, written on two corporations [13] [see also the discussions in Benzoni et al. (2012) and Bai et al. (2014)]. Recently Gourieroux et al. (2014) derived the pricing formulas for credit derivatives written on a large pool of corporations and taking into account the jumps arising when corporations in the pool default.

Finally formulas providing the prices of insurance contracts written on two lives depend on parameters explaining how the exogenous variable impact the bivariate lifetime (risk-neutral) distribution. These variables include the individual characteristics of the couple, in particular the information on their generation. This generation information for each given age allows for taking into account the deterministic time dependence of the mortality rate. Moreover, the unobserved explanatory variables can also depend on time in a stochastic way. Thus the longevity feature can be taken into account either by introducing generation (time) as an explanatory variable, or by introducing unobserved dynamic factor [see Duffie et al. (2009) for an example of unobserved dynamic Gaussian factor in credit risk modelling]. The parameters have to be calibrated, especially the parameters measuring the magnitude of the jumps (or of the association measures), the parameters capturing the frailty and how they depend on generation (i.e. time). We explain in Appendix A.4 why all the intensities are nonparametrically identified, in a mixed proportional hazard model, whenever the generation (cohort) effect is taken into account. The development of nonparametric, or semi-parametric, estimation methods is out of the scope of this paper on pricing, but they will clearly require enough data on coupled lives, disaggregated by generations of spouses and contracts.

13. which is equivalent to an insurance product written on two lives.

Appendix A.1 Joint density of lifetimes

Let us assume $y_1 < y_2$. We have :

$$\begin{aligned}
f(y_1, y_2) &= \lim_{dy_1, dy_2 \to 0} \frac{1}{dy_1 dy_2} P[Y_1 \in (y_1, y_1 + dy_1), Y_2 \in (y_2, y_2 + dy_2)] \\
&= \lim_{dy_1, dy_2 \to 0} \frac{1}{dy_1 dy_2} P[X_1 < X_2, X_1 \in (y_1, y_1 + dy_1), X_1 + X_4 \in (y_2, y_2 + dy_2)] \\
&= \lim_{dy_1, dy_2 \to 0} \left[\frac{1}{dy_1} P[y_1 < X_2, X_1 \in (y_1, y_1 + dy_1)] \right. \\
&\quad \left. \frac{1}{dy_2} P[X_4 \in (y_2 - y_1, y_2 - y_1 + dy_2) | X_1 = \min(X_1, X_2) = y_1] \right] \\
&= \left[-\frac{\partial S_{12}}{\partial x_1}(y_1, y_1) \right] \left[-\frac{\partial S_4}{\partial x_4}(y_2 - y_1; y_1) \right].
\end{aligned}$$

Appendix A.2 Link between the historical and risk-neutral distributions

For expository purpose we set the risk-free rate $r = 0$. Then we have to consider jointly the historical (or physical) distribution, with characteristics indexed by P, and the risk-neutral (or adjusted for risk) distribution, with characteristics indexed by Q. Since we are in an incomplete market frameworks, these two distributions can be specified independently. Let us now discuss the possible effects of the change of probability.

i) The stochastic discount factor (sdf) is the ratio between the risk-neutral and historical densities :

$$m(y_1, y_2, F) = \frac{f^Q(y_1, y_2, F)}{f^P(y_1, y_2, F)},$$

for a model with frailty for instance. A discontinuity of the risk-neutral density f^Q on the 45° line $y_1 = y_2$, that is, jumps in the risk-neutral intensities, can result from either jumps in the historical intensities, or jumps in the adjustment for risk (sdf) when a death occurs.

The standard insurance literature computing the prices from a specification of the historical distribution and the sdf has omitted the second possibility. This is typical of the practice of pricing by Esscher transforms [see e.g. Esscher (1932), Gerber and Shiu (1994)] written on factor F, that is choosing $m(y_1, y_2, F) = \exp(\alpha + \beta F)$, where α and β are such that $E^P[\exp(\alpha + \beta F)] = 1$

to get the zero risk-free rate.

Intuitively to reintroduce the effect of death event while using the practice of Esscher transforms, we may introduce the Esscher transforms on the distributions of the latent variables, that is,

$$\begin{aligned}
&\text{for the pair } (X_1, X_2) : \exp(\alpha_{12} + \beta_{12} F), \text{ say,} \\
&\text{for the pair } X_3 : \exp(\alpha_3 + \beta_3 F), \text{ say,} \\
&\text{for the pair } (X_4) : \exp(\alpha_4 + \beta_4 F), \text{ say.}
\end{aligned}$$

with parameters linked by the condition of zero risk-free rate.

Appendix A.3 Probability distribution function of the heterogeneity given survival up to time t

We derive the probability density function of the heterogeneity of the set of couples such that both spouses survive up to age $z_0 + x$. It is denoted g_x, We also denote by g_0 the heterogeneity distribution at age $z_0 = 30$, which equals $\gamma(k, 1/k)$, therefore :

$$g_0(f) \propto f^{k-1} \exp[-kf].$$

The unconditional survival probability that both survive up to age $z_0 + x$ is :

$$\begin{aligned}
S(x) &= \mathbb{P}(Y_1 > z_0 + x, Y_2 > z_0 + x | Y_1 > z_0, Y_1 > z_0) \\
&= \int \exp[-[A_1(x) + A_2(x)]f] g_0(f) \mathrm{d}f,
\end{aligned}$$

where A_1 and A_2 are cumulative intensities. Then the unconditional mortality intensity at age $z_0 + x$ is :

$$\begin{aligned}
\lambda(x) &= -\frac{\mathrm{d}}{\mathrm{d}x} \log S(x) \\
&= \frac{\int [a_1(x) + a_2(x)] f \exp[-[A_1(x) + A_2(x)]f] g_0(f) \mathrm{d}f}{\int \exp[-[A_1(x) + A_2(x)]f] g_0(f) \mathrm{d}f}.
\end{aligned}$$

Therefore, we deduce that the heterogeneity distribution function is :

$$g_x(f) = \frac{g_0(f)\exp[-[A_1(x)+A_2(x)]f]}{\int g_0(f)\exp[-[A_1(x)+A_2(x)]f]\mathrm{d}f}$$
$$\propto f^{k-1}\exp[-[k+A_1(x)+A_2(x)]f],$$

which is a gamma distribution with shape parameter k and scale parameter $1/(k+A_1(x)+A_2(x))$.

Appendix A.4 Identification of the model

To illustrate the possibility of nonparametric identification, let us consider a mixed proportional hazard model, where the latent intensities are of the type :

$$\lambda_j(t|z,x_j) = a_j(x_j)b_j(z)F_j, \qquad j=1,2,3,4, \tag{A-1}$$

where z are the observable individual covariates, F_j unobserved heterogeneity, a_j baseline intensities. The observed covariates can be the generation [14], as well as the date of the event $\min(Y_1,Y_2)$ for variables $j=3,4$ to allow for semi-Markov intensities.

We can distinguish different models based on the specification (A-1) according to the observed durations :

— The model $M_{1,2}$, if we observe (Y_1,Y_2).

— The model $M_{1|2}$, if we observe $(Y_1,Y_2\mathbb{1}_{Y_2<Y_1}) = (X_1+X_3\mathbb{1}_{X_2<X_1}, X_2\mathbb{1}_{X_2<X_1})$. In this model, the main duration variable of interest is Y_1 and Y_2 is observed only if it is smaller than Y_1.

— The model $M_{2|1}$, if we observe $(Y_2,Y_1\mathbb{1}_{Y_2<Y_1}) = (X_2+X_4\mathbb{1}_{X_1<X_2}, X_1\mathbb{1}_{X_1<X_2})$. In this model, the main duration variable of interest is Y_2 and Y_2 is observed only if it is smaller than Y_2.

— The model $M_{1\wedge 2}$, if we observe $(\min[Y_1,Y_2], \mathbb{1}_{Y_2<Y_1})$.

14. As we pointed out earlier in the paper, there are at least three generation effects, that are respectively the cohort of the husband, the cohort of the wife, and the year of inception of the contract.

These models are embedded in the following sequence :

$$M_{1\wedge 2} \subset \begin{matrix} M_{1|2} \\ M_{2|1} \end{matrix} \subset M_{1,2}$$

Model $M_{1\wedge 2}$ is commonly called competing risks model [see e.g. Abbring and van den Berg (2003a)] and is used in the analysis of mortality by causes. Model $M_{1|2}$ (resp. $M_{2|1}$) is called semi-competing risks model [see e.g. Xu et al. (2010)] in biostatistics or (survival) models with treatment effect in microeconometrics [see Abbring and van den Berg (2003b)]. For instance, model $M_{1|2}$ is a model for mortality of individual 1 subject to the death of 2 as treatment. Due to the sequence of embedded models, any function identifiable under $M_{1\wedge 2}$ (resp. $M_{1|2}$, $M_{2|1}$) is also identifiable under $M_{1|2}$ and $M_{2|1}$ (resp. $M_{1,2}$). This allows for applying Proposition 4 in (Abbring and van den Berg, 2003b), valid for the identification of treatment effects in duration models. Under mild conditions [15], we can, in Model $M_{1|2}$, identify nonparametrically functions [16]

$$a_1, a_2, b_1, b_2, a_3, b_3 \text{ and the joint distribution of } F_1, F_2, F_3.$$

In Model $M_{2|1}$, we can identify :

$$a_1, a_2, b_1, b_2, a_4, b_4 \text{ and the joint distribution of } F_1, F_2, F_4.$$

Thus under $M_{1,2}$ we can identify all functions a_j, b_j, $j = 1, 2, 3, 4$, as well as the 3-dimensional distributions of (F_1, F_2, F_3) and (F_1, F_2, F_4).

In practice, we often assume that $F_1 = F_3$, $F_2 = F_4$, where F_1 and F_2 can be dependent. Under this additional assumption on unobserved heterogeneities, Model $M_{1,2}$ is nonparametrically identified.

15. Roughly speaking, the observed covariate $b_j(z)$ should cover a non empty open set, that is, there should be sufficient covariate variation among different couples.

16. Whereas in the standard competing risks model $M_{1\wedge 2}$, we can nonparametrically identify a_1, a_2, b_1, b_2 and the joint distribution of F_1, F_2 under the same mild conditions [see Abbring and van den Berg (2003a), Proposition 2].

Chapitre III

Long-Term Care and Longevity

Abstract

The increase of the expected lifetime, that is the longevity phenomenon, is accompanied by an increase of the number of seniors with a severe loss of autonomy. Because of the significant costs of long-term care (LTC) facilities, it is important to analyze the time spent in LTC state, as well as the probability of entering into this state during its lifetime, and how they evolve jointly with longevity across the different cohorts. Our paper considers such questions, when lifetime data are available, but LTC data are either unavailable, or available on too short periods, or too aggregated, or unreliable, as it is frequently the case.

We specify joint structural models of LTC, mortality, and longevity, and explain why parameters of these models are identifiable from only the lifetime data under reasonable assumptions. More precisely, we model the potential entry into LTC as a latent state, which creates a dynamic unobserved heterogeneity in the population when only the lifetime is observed. The methodology is applied to the cohort mortality data of French males, first with a deterministic trend and then with a dynamic and stochastic common latent factor. Prediction formulas for the hypothetical date of entry into LTC or the time spent in this state are then provided and illustrated using the same data set.

Keywords : Longevity, Long-Term Care (LTC), Semi-Competing Risks, Treatment effect, Unobserved Heterogeneity, Dynamic Frailty, Partial Observability, Identification.

III.1 Introduction

The general increase of human lifetime, that is the longevity phenomenon, has been largely illustrated in the demographic and insurance literatures [see e.g. Lee and Carter (1992)]. In average we observe an increase of 3 months per annum of the life expectancy [see e.g. Oeppen and Vaupel (2002)]. This increase is accompanied by an increase of the number of old people who potentially need long-term care (LTC henceforth) [1], but also a decrease of the probability of entering into LTC at any given age [see e.g. Manton et al. (1998)], as well as a decrease of the mortality intensity for individuals in LTC *ceteris paribus* [2]. A person enters into LTC when he/she becomes unable to live independently, measured by the ability to do some special Activities of Daily Living (ADL). This entry into LTC state is in general irreversible and is accompanied by a huge increase of mortality intensity. Because of the significant costs of LTC facilities, it is important to analyze this probability of entry, the time spent in this state as well as how they evolve with longevity. Are they almost independent of the longevity feature or do they increase at a similar rate ? Our paper answers these questions, when the lifetime data are available, but the LTC data are either unavailable, or available on too short periods, or weakly reliable.

We introduce in this paper joint models of LTC and mortality, based on the intensity of entry into LTC state and on the mortality intensities. The model disentangles the mortality intensities according to the time spent in LTC state. Moreover we assume that these intensities depend on an unobservable dynamic factor (or dynamic frailty) with nonstationary features, able to capture the longevity phenomenon and its potential impact on both mortality and LTC. This longevity factor can be assumed deterministic, or stochastic.

Such a joint model would be simple to estimate if individual data on both mortality and LTC were available [see e.g. Levantesi and Menzietti (2012), Majer et al. (2013)]. However data on LTC are often missing or not very reliable when they exist. Indeed, there does not even exist a universal definition of the LTC state. In the literature, the very terminology is often confounded [3] with “losing autonomy", “disability", “morbidity" or “nursing/home care" and differs by both country and insurance company ; further more, it is subject to changes across time. In the US, insurers consider six limitations of Activities of Daily Livings, that are Eating, Dressing, Walking, Bathing, Toileting, and Maintaining Continence, respectively, while their European

1. Also called nursing care in the literature.
2. That is, when all other parameters, for instance the current age, as well as the age of entry are equal.
3. For instance, Levantesi and Menzietti (2012) propose to price private LTC contracts using national disability benefit data.

peers, use only four of them called Instrumental Activities of Daily Living (IADL) [see e.g. Rice (1989), Kessler (2008) for a review of the LTC insurance market]. This discrepancy is even larger between public LTC insurance plans in different countries (often Western European), where it is a pillar of the social security system. An OECD disability indicator even include extra criteria such as hearing and reading small letters [see McWhinnie (1980)] ; French public databases based on different population samples show different trends of the LTC/disability prevalence [4] [see Lafortune and Balestat (2007)]. Finally, current data often measure the actual LTC use, instead of the need of LTC. There are various reasons for the two to differ in practice, such as administrative delay [5], the lack of self-diagnosis capacity of the disabled, or budget constraint, or even the incentive of false claim [6].

Moreover, even when data exist, they often lack accuracy. Indeed, collecting LTC data is a much more demanding task than collecting mortality data since it requires the knowledge of the entire history of each individual, especially the time(s) at which an IADL is lost, identified by accredited physicians. Most of the time, available public data of the national population only exist for a few years when there is either a census, or a sample population survey [7] with a large time spell between neighbouring surveys ; their quality are quite limited because of the voluntary nature of the survey responses and the fact that surveys conducted in different years do not necessarily concern the same individuals. Another problem is that most datasets are cross-sectional, either by nature, or because the observation period is too short to deliver longitudinal information. So from the very beginning they are not suited for the understanding of the evolution of the LTC risk. Indeed, by using such a cross-sectional database one will in general ignore the evolution in cohort of the different transition probabilities at given ages [see Keiding (1991) for a discussion on the limits of this stationary approach] ; this is unrealistic and dangerous given the potentially large impact of the longevity on both LTC and mortality risks. This uncertainty on the future evolution and its poor understanding is a serious obstacle to the further development

4. That is, the proportion of people in LTC.

5. For instance, it is common practice for insurance companies to acknowledge the entry into LTC of a policy-holder (and begin periodic benefit payment) only six months after the effective entry, to make sure that the entry is really permanent.

6. For instance, Dienst (1972) states that during past severe economic crisis, the number of people declaring disabled increased. This effect is produced mainly by people who have been medically disabled long time ago and in addition by people with relatively minor medical problems who would not consider themselves disabled in good times, but who in both instances are induced to claim insurance benefits only in case of a crisis.

7. This is for instance the case for the survey “Handicaps-Incapacité-Dépendances" in France (literally the Disability-Incapacity-Long-Term Care Survey, this survey has been conducted in 1998/1999 and then in 2008/2009.), as well as the National Long-Term Care Survey (NLTCS) database in the US (which is based on surveys conducted in 1982,1984,1989,1994,1999, 2004 on a representative sample of the US population, see its official website `http://www.nltcs.aas.duke.edu/`). These two countries are also by far the two largest markets for private LTC insurance.

of the private LTC insurance market in many countries, in a period when the sustainability of the Welfare States is more and more questioned and the public's appetite for private LTC insurance is steadily increasing.

Our paper develops a methodology to estimate this joint model of risks using only the mortality data. Rather than relying on data with an *ad hoc* definition of the LTC state, we consider the autonomy state as a latent state variable and the entry into LTC is characterized by an unobservable mortality jump [8]. The assumption that we can capture an individual's aging history by such a model with two regimes, and interpret one of them as the entry into LTC is not just for identification convenience. Indeed, physiologically speaking, the entry into LTC is not an independent event, but is often caused by random events such as the onset of a disease or an accident [9]. Not all such events result in LTC, which becomes necessary only when there is a significant deterioration of the health, accompanied by a major rise of the mortality intensity. This change of regime is by nature latent, and is only imperfectly captured by existing data on LTC. Our model provides an "optimal" definition of LTC which which we will "filter" out of the lifetime data.

Due to the higher mortality for people in LTC, when the mortality is analyzed using only lifetime data, the autonomy state at a given age [10] is a time-dependent unobserved heterogeneity. Therefore there is a spurious duration dependence as in a population with static unobserved heterogeneity, or static frailty [see e.g. Vaupel et al. (1979) and Elbers and Ridder (1982)]. This effect should be identified in order to study the true duration dependence, that is, the age dependence of the mortality evolution, and how this dynamics changes between different cohorts, that is the longevity phenomenon. Under reasonable assumptions, the possibility to identify the characteristics of LTC from the mortality data is due to the jumps in mortality intensity arising when entering into LTC and to the assumed effects of the unobserved longevity factor on both mortality and LTC across different cohorts. Thus, such a model allows us to predict jointly the future evolution of the LTC entry probabilities and the mortality intensities.

The paper is organized as follows. In Section 2, we introduce a joint modeling of LTC and mortality risks. This modeling is used in Section 3 to derive the joint distribution of the lifetime

8. The idea of introducing latent state variables is recently also proposed by Wouterse et al. (2013). With observations of a large number of health indicators including the LTC status, they construct a latent state variable as a synthetic measure of the individual's health status. However, in their framework, LTC is observable and their methodology does not allow for an analysis of the evolution of various risks across different cohorts.

9. For instance, Kessler (2008) claims that more than 70 % of LTC entries is caused by chronic diseases such as cancer and dementia, others being triggered by events such as accidents or mental diseases.

10. Either autonomous, or in LTC.

and of the date of entry into LTC. To derive this distribution we follow a progressive approach. We first consider the case of “observable" intensities, then we render them stochastic by introducing a static frailty. In Section 3 we consider a basic model with constant intensities and discuss its identification. Section 4 introduces semi-parametric specifications for the intensities and the frailty dynamics, discuss the way of introducing a nonstationary longevity generation effect, solve the identification issues, and derive the form of the log-likelihood function when the lifetimes are observed with right censoring. The models are estimated for the French male population in Section 5. We first consider a model with deterministic factor in the spirit of the Lee-Carter model, but allowing for non degenerate intensities in a far future. We allow for either Markov or semi-Markov mortality intensity functions. Then the model is extended to include the uncertainty on the longevity factor by means of a dynamic frailty process. We also explain how to filter out this frailty process once the model is estimated. In Section 6 we implement the model for prediction purpose. Section 7 concludes. Proofs and other technical details are gathered in Appendices.

III.2 Structural versus reduced form approach

Let us consider a situation where an individual can either experience first a **non terminal event** and then fail, or can fail directly. In both situations the failure is called the **terminal event**. In the second case, the terminal event censors the non terminal event. The corresponding model is called semi-competing risks [11] in the literature [see e.g. Fine et al. (2001), Xu et al. (2010)]. In our framework, the non terminal event is the potential entering into LTC and the terminal event is the death. The migration from the autonomous state to the LTC is assumed irreversible. Thus there is an asymmetry between both types of events.

We first introduce a structural approach with latent variables corresponding to the times elapsed up to the potential events and describe how the ideally observable variables depend on the latent duration variables. Then we derive an alternative methodology in terms of intensities.

In the literature, most multivariate survival models are written in continuous time. The main reason is that in the continuous time intensity-based setting, the probability of observing tied events is naturally null. In our example, we would like to avoid the simultaneous arrival of both the non terminal and the terminal event. Thus we follow the continuous time approach, at least

11. In the microeconometric literature, the effect of the non terminal event on the terminal event is also called "treatment effect" [see e.g. Abbring and van den Berg (2003b)], even if the exogenous entry in LTC cannot really be interpreted as a treatment as in other types of economic applications.

for the theoretical model. The continuous time model is discretized when it comes to numerical estimation of the model with dynamic frailty.

We begin our analysis by considering only one cohort (generation). In this case and without left censoring (which we also assume for the time being), we can use either the terminology “age" or “time" to denote the elapsed duration. From Section 4 on, when the cohort effect is introduced, we will more frequently use the term “age" for the elapsed duration, that is, the age of an individual since its birth. To describe the period effect, we use the term “calendar time" and we have the following relationship between the three time measures :

$$\text{Cohort birth date} + \text{Age} = \text{Calendar time}.$$

III.2.1 Structural approach

Semi-competing risks are traditionally written on the two duration variables Y_1^* and Y_2, where Y_2 is the failure time and Y_1^* is the potential time of entering into LTC. Therefore, the variable Y_1^* is latent since it is not observable when we observe first the variable Y_2, that is, when $Y_2 < Y_1^*$. Then the dependence between the two variables is often modeled via a survivor copula C [see e.g. Fine et al. (2001) and Hsieh et al. (2008)], that is,

$$\mathbb{P}(Y_1^* > y_1, Y_2 > y_2) = C(S_1(y_1), S_2(y_2)), \tag{III-1}$$

where C is assumed to belong to some specific parametric families, e.g. Archimedean copulas or other factor copulas and S_1, S_2 denote the marginal survivor functions of Y_1^* and Y_2, respectively. This bivariate copula approach is partly borrowed from the literature on competing risks models [see e.g. Zheng and Klein (1995)]. The model is often written with restrictions such as a continuous copula density, and a positive, symmetric dependence structure. But such a direct modeling is not flexible enough to capture the peculiarities of semi-competing risks data. First, they are not adapted to characterize the “regime switching" nature that an individual may experience. Intuitively, if the individual enters into the LTC during its lifetime, then his residual lifetime distribution will be very different from the case when he never experiences the LTC. Therefore, using solely one variable Y_2 to model the lifetime is probably not enough. Besides, the idea behind equation (III-1) is that instead of being latent, the variable Y_1^* is treated as observable (and is only censored when $Y_2 < Y_1^*$ instead of being nonexistant). This confusion explains also

the decades-long debate on the physical meaning of the latent variables in (semi)-competing risks models [see Prentice et al. (1978) and Andersen and Keiding (2012)]. We consider below an alternative approach with an extra latent variable. More precisely, let us introduce :

— X_1 the potential time of entry in LTC,

— X_2 the (potential) time of death for an individual which has not experienced LTC,

— X_3 the residual lifetime up to the death once the individual experienced LTC.

Some of these variables are really latent even for an econometrician with the maximal available information. Indeed an individual dying before the potential entry in LTC will never experience spell X_1, or X_3. At most the observations include the indicator variable Z defined by : $Z = \mathbb{1}_{X_1 \leq X_2}$, that is, whether or not the individual experiences the LTC before the death, and the duration variable(s) :

$$\begin{cases} Y_1^* = X_1 \text{ and } Y_2 = X_1 + X_3, & \text{if } Z = 1, \\ Y_2 = X_2, & \text{if } Z = 0. \end{cases} \tag{III-2}$$

In regime 1, we ideally observe the time Y_1^* up to the entry into LTC and the lifetime Y_2. In regime 0, we observe the lifetime only.

The ideally observable model can be rewritten in another form, which avoids the explicit distinction between the regimes. For this purpose, we introduce a variable Y_1 defined by $Y_1 = Y_1^*$, if $Z = 1$, and $Y_1 = 0$, otherwise, which captures both the regime and the duration up to the non terminal event, if the latter is observed. We get :

$$\begin{cases} Y_1 &= X_1 Z, \\ Y_2 &= (X_1 + X_3)Z + X_2(1 - Z). \end{cases} \tag{III-3}$$

The first equation corresponds to a standard Tobit model [see e.g. Amemiya (1984)] and is completed by an equation providing the observed lifetime depending on the regime.

To our best knowledge, the idea of introducing explicitly a regime change dates back to Freund (1961), who considered only the case of constant hazards ; it is later generalized to the previous general form by Tosch and Holmes (1980). Recently this latent model has been generalized to include static frailty [see Abbring and van den Berg (2003b)] and an extended version applied to the pricing of joint insurance contracts for couples [see Gouriéroux and Lu (2013)]. The aim of our paper is to introduce dynamic (common) frailty featuring trends and able to capture the

stochastic longevity phenomenon.

In general, latent variables X_1, X_2, X_3 are specified by means of their hazard functions as well as some assumptions on the dependence between them. The next subsection gives a natural interpretation of these hazard functions in terms of transition intensities of an individual between different health states.

III.2.2 Reduced form approach

The model can also be defined by a chain with the three following states :

— state A : the individual is autonomous,

— state B : the individual is under LTC,

— state C : the individual is dead. State C is the unique absorbing state.

The transitions are possible only from state A to state B, from state B to state C and from state A to state C. The history of the individual is represented by the qualitative process $S = (S_t)$ which takes value in the state space $\{A, B, C\}$. The scheme below gives the possible paths of an individual's lifetime.

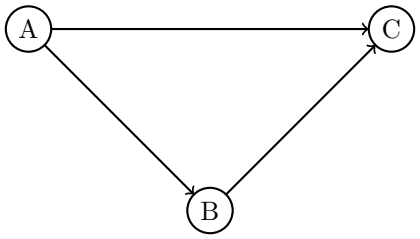


FIGURE III-1: The potential transitions of an individual during its lifetime.

Let us denote by $\underline{S_t}$ the information on past individual history up to time t : $\underline{S_t} = \{S_u, 0 \leq u \leq t\}$, then we define the following transition intensities :

$$\text{If } S_t = A, \mu_1(t) = \lim_{du \to 0^+} \Big\{ \frac{1}{du} \mathbb{P}(S_{t+du} = B | \underline{S_t}) \Big\},$$
$$\text{If } S_t = A, \mu_2(t) = \lim_{du \to 0^+} \Big\{ \frac{1}{du} \mathbb{P}(S_{t+du} = C | \underline{S_t}) \Big\},$$
$$\text{If } S_s = S_t = B, S_{s-} = A, \mu_3(t|s) = \lim_{du \to 0^+} \Big\{ \frac{1}{du} \mathbb{P}(S_{t+du} = C | \underline{S_t}) \Big\}, \qquad \forall t > s.$$

Due to the qualitative nature of process (S_t), the knowledge of $\underline{S_t}$ is equivalent to the knowledge of its current state, of its previous state (if it exists) and of the corresponding transition time.

Therefore we can rewrite the transition intensities as follows :

$$\mu_1(t) = \lim_{du \to 0^+} \Big\{ \frac{1}{du} \mathbb{P}(S_{t+du} = B | S_t = A) \Big\},$$
$$\mu_2(t) = \lim_{du \to 0^+} \Big\{ \frac{1}{du} \mathbb{P}(S_{t+du} = C | S_t = A) \Big\},$$
$$\mu_3(t|s) = \lim_{du \to 0^+} \Big\{ \frac{1}{du} \mathbb{P}(S_{t+du} = C | S_{s-} = A, S_s = S_t = B) \Big\}.$$

The conditions on intensities μ_1 and μ_2 are Markov conditions. The condition on μ_3 is a semi-Markov condition since the transition also depends on the time of entry into LTC. This reduced form approach is more commonly called the illness-death model. Its usefulness in modeling semi-competing risks has only been rediscovered recently by Xu et al. (2010).

It is easily checked that (see Section 3.1) this reduced form specification is equivalent [12] to the structural model we defined in Section 2.1, if we carefully specify the intensity functions of the latent variables and the dependence structures between them. This should diminish the considerable confusion in the literature that the reduced form approach is different from the structural approach and that it should be preferred [see e.g. Imai and Soneji (2007)]. However, in some applications, one approach may be more convenient than the other one. To quote a summary from Han and Hausman (1990) : "While econometricians have emphasized the presence of unobserved heterogeneity" (and therefore prefer the structural approach), "statisticians have instead emphasized the use of semi-parametric models which do not require parametric specification of the baseline hazard" (hence the choice of reduced form approach, often written without unobserved heterogeneity).

III.3 The distribution of the potentially observable variables

Let us now derive the explicit expressions of the joint distribution of variables (Y_1, Y_2), and also of the marginal distribution of Y_2. We consider the case in which the latent variables X_1, X_2 are independent. Then we discuss the structural model with constant intensities to highlight the identification issues.

12. The only difference is that in the latent variable approach, the variable X_3 is defined even if $X_1 > X_3$. But in such cases the value of X_3 is not important.

III.3.1 The basic model

Joint distribution of the latent variables

Let us first assume that the latent variables X_1, X_2 are independent. Their joint distribution is characterized by their marginal intensities :

$$\lambda_1(x_1) = \lim_{du \to 0^+} \Big\{ \frac{1}{du} \mathbb{P}(X_1 \leq x_1 + du | X_1 \geq x_1) \Big\},$$

$$\lambda_2(x_2) = \lim_{du \to 0^+} \Big\{ \frac{1}{du} \mathbb{P}(X_2 \leq x_2 + du | X_2 \geq x_2) \Big\}.$$

The variable X_3 is in general defined conditional on the values of X_1 and X_2, and is often assumed independent of X_2. Therefore we denote by $\lambda_{2|1}(x_3|x_1)$ its intensity given the value of $X_1 = x_1$, which depends both on the non terminal event time x_1 and the time elapsed since the non terminal event x_3. :

$$\lambda_{2|1}(x_3|x_1) = \lim_{du \to 0^+} \Big\{ \frac{1}{du} \mathbb{P}(x_3 \leq X_3 + du | X_3 > x_3, X_1 = x_1) \Big\}.$$

When this function depends on x_1, x_3 only via $x_1 + x_3$, the model is Markov ; otherwise, it is semi-Markov.

The joint density function of the latent variables (X_1, X_2, X_3) is :

$$g(x_1, x_2, x_3) = e^{-\Lambda_1(x_1) - \Lambda_2(x_2) - \Lambda_{2|1}(x_3|x_1)} \lambda_1(x_1) \lambda_2(x_2) \lambda_{2|1}(x_3|x_1),$$

where $\Lambda_1, \Lambda_2, \Lambda_{2|1}$ are the cumulated intensities associated with $\lambda_1, \lambda_2, \lambda_{2|1}$, respectively. Therefore the joint survival function of the latent variables (X_1, X_2, X_3) is :

$$\begin{aligned} S(x_1, x_2, x_3) &= \int_{x_1}^{\infty} \int_{x_2}^{\infty} \int_{x_3}^{\infty} e^{-\Lambda_1(t_1) - \Lambda_2(t_2) - \Lambda_{2|1}(t_3|t_1)} \lambda_1(t_1) \lambda_2(t_2) \lambda_{2|1}(t_3|t_1) dt_1 dt_2 dt_3 \\ &= e^{-\Lambda_2(x_2)} \int_{x_1}^{\infty} \int_{x_3}^{\infty} e^{-\Lambda_1(t_1) - \Lambda_{2|1}(t_3|t_1)} \lambda_1(t_1) \lambda_{2|1}(t_3|t_1) dt_1 dt_3 \\ &= e^{-\Lambda_2(x_2)} \int_{x_1}^{\infty} e^{-\Lambda_1(t_1) - \Lambda_{2|1}(x_3|t_1)} \lambda_1(t_1) dt_1. \end{aligned}$$

Under these independence assumptions, we get :

$$\begin{aligned}
&\text{If } S_t = A, \qquad \mu_1(t) = -\frac{\partial}{\partial y_1} \log S_{12}(t,t) = \lambda_1(t),\\
&\text{If } S_t = A, \qquad \mu_2(t) = -\frac{\partial}{\partial y_2} \log S_{12}(t,t) = \lambda_2(t),\\
&\text{If } S_s = S_t = B, S_{s-} = A, \qquad \mu_3(t|s) = \lambda_{2|1}(t-s|s), \qquad \forall t > s,
\end{aligned}$$

where S_{12} is the joint survivor function $S_{12}(t_1, t_2) = \mathbb{P}[X_1 > t_1, X_2 > t_2]$. Therefore the structural approach with latent variables is equivalent to the reduced form approach. This equivalence is easily extended when (possibly unobserved and/or time-varying) stochastic factors are introduced, if we assume that (X_1, X_3) and X_2 are independent given the whole history of the factors and we define the transition intensities conditional on the whole history of the factors. The rest of the paper will use the structural approach, but keeping in mind this equivalence can certainly help the reader better understand certain formulas.

Distribution of the ideally observable variables

Let us now derive the joint distribution of the ideally observable variables (Y_1, Y_2). The couple (Y_1, Y_2) has a bi-dimensional continuous component on domain $\mathcal{D}_1 = \{(y_1, y_2) : y_1 < y_2\}$, and a one-dimensional continuous component on $\mathcal{D}_0 = \{(y_1, y_2) : y_1 = 0, y_2 > 0\}$. The joint distribution of (Y_1, Y_2) admits a density with respect to the dominating measure $\lambda_{\mathcal{D}_1} + \lambda_{\mathcal{D}_0}$, where $\lambda_{\mathcal{D}}$ denotes the Lebesgue measure on domain $\mathcal{D}$. This density is :

$$f(y_1, y_2) = \lambda_1(y_1)\lambda_{2|1}(y_2 - y_1|y_1)e^{-\Lambda_1(y_1) - \Lambda_2(y_1) - \Lambda_{2|1}(y_2 - y_1|y_1)}, \text{ on domain } \mathcal{D}_1 = \{(y_1, y_2) : y_1 < y_2\}, \tag{III-4}$$

and

$$f(0, y_2) = \lambda_2(y_2)e^{-\Lambda_1(y_2) - \Lambda_2(y_2)}, \text{ on domain } \mathcal{D}_0 = \{(y_1, y_2) : y_1 = 0, y_2 > 0\}. \tag{III-5}$$

Many authors write instead the joint distribution of (X_1, Y_2) [see also Xu et al. (2010) for a discussion], in which case there will be no point mass, but instead a continuous component on the unobservable domain $\{X_1 > Y_2\}$ and the restriction of the density function adds up to $\mathbb{P}[X_1 > Y_2] = \mathbb{P}[Y_1 = 0]$ there. These two approaches are equivalent, since in any application the latent variable should be integrated out. Nevertheless, as explained at the beginning of Section

2.1, studying directly (Y_1, Y_2) is preferred in order to distinguish explicitly the ideally observable information, that is (Y_1, Y_2), from the really latent one (X_1, X_2, X_3).

We deduce the marginal survival function and the p.d.f. of the lifetime Y_2, which is later on the only really observable duration variable :

Property III.1. The survival function of the lifetime Y_2 is :

$$S_2(y_2) = \mathbb{P}(Y_2 > y_2) = \int_0^{y_2} \lambda_1(t) e^{-\Lambda_1(t) - \Lambda_2(t) - \Lambda_{2|1}(y_2 - t|t)} dt + e^{-\Lambda_1(y_2) - \Lambda_2(y_2)}, \tag{III-6}$$

and its p.d.f. is :

$$f_2(y_2) = \int_0^{y_2} \lambda_1(t) \lambda_{2|1}(y_2 - t|t) e^{-\Lambda_1(t) - \Lambda_2(t) - \Lambda_{2|1}(y_2 - t|t)} dt + \lambda_2(y_2) e^{-\Lambda_1(y_2) - \Lambda_2(y_2)}. \tag{III-7}$$

Proof : See Appendix B.1. □

III.3.2 Identification in a model with constant intensities

For illustration purpose, let us assume a model with constant intensities λ_1, λ_2, and $\lambda_{2|1}$, that is with independent exponential latent variables. This simplified framework is useful to highlight the identification issue when only the lifetime variable Y_2 is observed.

For constant intensities the joint density becomes :

$$f(y_1, y_2) = \lambda_1 \lambda_{2|1} e^{-\lambda_1 y_1 - \lambda_2 y_1 - \lambda_{2|1}(y_2 - y_1)}, \text{ on domain } \mathcal{D}_1 = \{(y_1, y_2) : y_1 < y_2\},$$

and

$$f(y_1, y_2) = \lambda_2 e^{-(\lambda_1 + \lambda_2) y_2}, \text{ on domain } \mathcal{D}_0 = \{(y_1, y_2) : y_1 = 0, y_2 > 0\}.$$

The marginal survivor function of lifetime Y_2 becomes :

$$\begin{aligned} S_2(y_2) = \frac{\lambda_1}{\lambda_1 + \lambda_2} \Big[\frac{\lambda_1 + \lambda_2}{\lambda_1 + \lambda_2 - \lambda_{2|1}} e^{-\lambda_{2|1} y_2} - \frac{\lambda_{2|1}}{\lambda_1 + \lambda_2 - \lambda_{2|1}} e^{-(\lambda_1 + \lambda_2) y_2} \Big] \\ + \frac{\lambda_2}{\lambda_1 + \lambda_2} e^{-(\lambda_1 + \lambda_2) y_2}, \text{ if } \lambda_1 + \lambda_2 \neq \lambda_{2|1}, \end{aligned} \tag{III-8}$$

and

$$S_2(y_2) = \frac{\lambda_1}{\lambda_1+\lambda_2}\Big[1+(\lambda_1+\lambda_2)y_2\Big]e^{-(\lambda_1+\lambda_2)y_2} + \frac{\lambda_2}{\lambda_1+\lambda_2}e^{-(\lambda_1+\lambda_2)y_2}, \text{ if } \lambda_1+\lambda_2=\lambda_{2|1}. \tag{III-9}$$

Both functions :

$$y \mapsto \frac{\lambda_1+\lambda_2}{\lambda_1+\lambda_2-\lambda_{2|1}}e^{-\lambda_{2|1}y} - \frac{\lambda_{2|1}}{\lambda_1+\lambda_2-\lambda_{2|1}}e^{-(\lambda_1+\lambda_2)y},$$

and

$$y \mapsto \Big[1+(\lambda_1+\lambda_2)y\Big]e^{-(\lambda_1+\lambda_2)y},$$

are survivor functions (see Appendix B.2). In both cases ($\lambda_1+\lambda_2-\lambda_{2|1}=0$, or $\neq 0$), the distribution of lifetime Y_2 is a mixture of an exponential distribution with parameter $\lambda_1+\lambda_2$, and a gamma distribution, $\gamma(2,\lambda_1+\lambda_2)$, when $\lambda_{2|1}=\lambda_1+\lambda_2$. This decomposition has the following interpretation :

$$\mathbb{P}(Y_2>t) = \mathbb{P}(Z=0)\mathbb{P}(Y_2>t|Z=0) + \mathbb{P}(Z=1)\mathbb{P}(Y_2>t|Z=1),$$

with $\mathbb{P}(Z=1)=\mathbb{P}(X_1<X_2)=\frac{\lambda_1}{\lambda_1+\lambda_2}$.

Let us now discuss the identification of all parameters including the parameter $\lambda_{2|1}$ driving the time spent in LTC, when only the lifetime is observed. The following Proposition is a consequence of equations (III-8) and (III-9) :

Property III.2. Consider the model with constant intensities and assume that the lifetime Y_2 is the only observable variable.

i) If $\lambda_1+\lambda_2-\lambda_{2|1}\neq 0$ and $\lambda_2\neq\lambda_{2|1}$,

the mixture representation has two distinct components and the three parameters $\lambda_1,\lambda_2,\lambda_{2|1}$ can be identified from the distribution of lifetime Y_2 given in equation (III-8).

ii) If $\lambda_2=\lambda_{2|1}$,

the non terminal event has no effect on the mortality intensity. We get $S_2(y_2)=e^{-\lambda_{2|1}y}$. The parameter $\lambda_2=\lambda_{2|1}$ is identifiable, but not the parameter λ_1.

iii) If $\lambda_1+\lambda_2-\lambda_{2|1}=0$,

the expression of $S_2(y_2)$ is given by equation (III-9), and the three parameters $\lambda_1,\lambda_2,\lambda_{2|1}$

can all be identified.

Therefore, under the assumption of constant intensities, the possibility of identifying the parameters is based on the jump in mortality intensity upon entry into LTC, that is, on the regime switch. Such a jump exists if and only if the point process associated with the LTC state causes the point process corresponding to mortality [see e.g. Abbring and van den Berg (2003b)].

However, Proposition 2 *iii*) has to be interpreted carefully. The three parameters are identifiable, only if it is known *ex-ante* that the constraint $\lambda_1 + \lambda_2 - \lambda_{2|1} = 0$ is satisfied.

III.4 Model with longevity effect

III.4.1 An identification issue

The model with constant intensity is not appropriate for modeling longevity effects in lifetime and LTC analysis. The longevity factor can be represented by introducing in the latent intensities a positive variable F indexed by calendar time. More precisely, let us consider a generation of individuals indexed by the birth date t_0, that is, the (stochastic) calendar date of death of an individual of this generation is $t_0 + Y_2$. The three intensities given the whole history $\underline{F}$ of the longevity factor are of the following form :

$$\begin{cases} \lambda_1(x_1|\underline{F}, t_0) & = \lambda_1(x_1, F_{t_0}) = a_1(x_1) + b_1(x_1)F_{t_0+x_1}, \\ \lambda_1(x_2|\underline{F}, t_0) & = \lambda_2(x_2, F_{t_0}) = a_2(x_2) + b_2(x_2)F_{t_0+x_1}, \\ \lambda_{2|1}(x_3|\underline{F}, x_1, t_0) & = \lambda_{2|1}(x_3|x_1, F_{t_0}) = a_3(x_3|x_1) + b_3(x_3|x_1)F_{t_0+x_1+x_3}. \end{cases} \tag{III-10}$$

where $a_1(\cdot), a_2(\cdot)$,$a_3(\cdot|\cdot)$, $b_1(\cdot), b_2(\cdot)$,$b_3(\cdot|\cdot)$ are positive (hazard) functions.

The specification (III-10) disentangles the effect of age and of the current date in the intensities. The longevity factor is introduced as usual in a linear way. Since the factor is expected with a (deterministic or stochastic) trend, the linearity assumption implies cointegration between the different intensities with cointegrating vectors depending on age. This cointegration feature is introduced to capture the extension of lifespan going hand in hand with an extension or a diminution (according to the countries) of the amount of life spent in LTC. To get interpretable intensities for any generation, especially when t_0 tends to infinity, we consider a trend effect such that $\lim_{t\to\infty} F_t = 0$. Under this condition, when t_0 goes to infinity, the intensities converge to $a_1(x_1)$, $a_2(x_2)$ and $a_3(x_3|x_1)$, respectively. Thus these functions can be interpreted as long term

intensities, that are intensities in a far future. This is one difference with the basic Lee-Carter model [Lee and Carter (1992)] where in a far future the intensities are assumed equal to zero, that is, where the individual will necessarily become eternal.

Model (III-10) is semi-parametric with unknown functions $a_1(x_1)$, $a_2(x_2)$, $a_3(x_3|x_1)$, $b_1(x_1)$, $b_2(x_2)$, $b_3(x_3|x_1)$, and the dynamics of the longevity factor, which will be parameterized in the next subsection. This is a constrained structural model, but these constraints are not sufficient to identify all unknown parameters from just the observation of the lifetime Y_2, even if we have jump in the intensities and the generation can be viewed as a covariate. Indeed, in the limiting case when the generations have infinite sizes and all generations are observed, the observable distribution summary is the survivor function indexed by the generation $S_2(y_2; t_0)$ [see Equation (III-13) for a typical expression of this function]. This is a function on $]0, \infty[^2$, but the set of functions to be estimated already includes two functions $a_3(x_3|x_1)$ and $b_3(x_3|x_1)$ defined on the same space. Then the order condition for identification is not satisfied. Such a lack of identification is standard in models with treatment effects [see e.g. Abbring and van den Berg (2003b)]. It is here observed despite restrictions already introduced on the models and the effect of two exogenous variables, i.e., the observed indicator of the cohort and the unobserved longevity factor.

Thus to recover the identification of the joint distribution of the latent intensities (X_1, X_2, X_3), we need additional restrictions. We will assume that the conditional intensities $a_3(x_3|x_1)$ and $b_3(x_3|x_1)$ can be written in terms of univariate functions defined on $]0, \infty[$.

III.4.2 Constrained specifications

In the application we will consider two constrained specifications.

Specification of the baseline intensities

The first specification corresponds to the Markov case, where the intensity $\lambda_{2|1}(x_3|x_1, t_0)$ depends on x_3 and x_1 through the current age $x_3 + x_1$ only :

$$\begin{cases} a_3(x_3|x_1) & = a_3(x_3 + x_1), \\ \tilde{b_3}(x_3|x_1) & = \tilde{b_3}(x_3 + x_1). \end{cases} \tag{III-11}$$

We will also consider the following semi-Markov model,

$$\begin{cases} a_3(x_3|x_1) & = a_4(x_3) + a_5(x_1), \\ \tilde{b_3}(x_3|x_1) & = b_4(x_3) + b_5(x_1) \end{cases} \tag{III-12}$$

with additive decomposition of the conditional intensities. For instance, under the Markov model (III-11), the survivor function of the observed variable y_2 given the future factor path $\overline{F}_{t_0} = \{F_\tau, \tau \geq t_0\}$ is :

$$\begin{aligned} S_2(y_2, t_0) = \int_0^{y_2} [a_1(x) + b_1(x)F_{t_0+x}] \exp\Big(- \int_0^x [a_1(s) + b_1(s)F_{t_0+s}]ds \\ - \int_0^x [a_2(s) + b_2(s)F_{t_0+s}]ds - \int_x^{y_{2,i}} [a_3(s) + b_3(s)F_{t_0+s}]ds\Big)dx \\ + \exp\Big(- \int_0^{y_2} [a_1(x) + b_1(x)F_{t_0+x}]dx - \int_0^{y_2} [a_2(x) + b_2(x)F_{t_0+x}]dx\Big). \end{aligned} \tag{III-13}$$

Specification of the factor dynamics

i) **Deterministic factor.** Let us first assume a deterministic factor (F_t), with exponential pattern :

$$F_t = \exp(-mt), \tag{III-14}$$

where $m > 0$. The factor is known up to the value of the parameter m.

Under the exponential specification (III-14), the age-calendar time model (III-10) can be equivalently written as an affine age-cohort model [13] :

$$\begin{cases} \lambda_1(x_1|\underline{F}, t_0) & = \lambda_1(x_1, F_{t_0}) = a_1(x_1) + \tilde{b}_1(x_1)F_{t_0}, \\ \lambda_2(x_2|\underline{F}, t_0) & = \lambda_2(x_2, F_{t_0}) = a_2(x_2) + \tilde{b}_2(x_2)F_{t_0}, \\ \lambda_{2|1}(x_3|\underline{F}, x_1, t_0) & = \lambda_{2|1}(x_3|x_1, F_{t_0}) = a_3(x_3|x_1) + \tilde{b}_3(x_3|x_1)F_{t_0}. \end{cases} \tag{III-15}$$

with, say, $\tilde{b}_1(x_j) = b_j(x_j)e^{-mx}, j = 1, 2, \tilde{b}_3(x_3|x_1) = b_3(x_3|x_1)e^{-mx_1 - mx_3}$.

In the age-calendar time model, the shocks on the factors depend on date t, whereas in the age-cohort model the factor has an impact at birth with consequences during the whole cohort lifetime. Thus, for exponential factor, it is not possible to distinguish between both interpretations of longevity, that is to say if longevity is associated with time, or with generation [see also

13. It is only in this exponential case that we have both an affine age-cohort model and an equivalent affine age-period model. Indeed, if we have both $\lambda_1(x_1|\underline{F}, t_0) = a_1(x_1) + b_1(x_1)F_{t_0+x_1}$ and $\lambda_1(x_1|\underline{F}, t_0) = \tilde{a}_1(x_1) + \tilde{b}_1(x_1)F_{t_0}$, it is easily shown that, given continuity assumptions on the function $t \mapsto F_t$, this function is necessarily an exponential function of time t.

Heckman and Robb (1985)].

The affine age-cohort specification is very similar to the popular proportional hazard models in survival analysis, in which the effect of the exogenous covariates, here the cohort t_0, appears often in a multiplicative way in the conditional intensity given the covariate. This model is mathematically easier to handle for nonparametric identification (see Appendix B.7). The coefficient $\tilde{b}_j, j = 1, 2, 3$ measure the persistence of different intensities with respect to the generation effect F_{t_0}.

However, the age-calendar time specification is also widely used in demography and finance. It assumes that the longevity phenomenon is instead more influenced by calendar year fluctuations which incorporates, besides a general decrease of mortality (due to e.g. the progress in medicine), temporary effects such as pandemic, natural disasters, etc. The nonparametric identification of an age-calendar time model, with an unconstrained F, is more difficult to study. Indeed, for a same cohort t_0, the intensity of the observed variable y_2 depends on the age x via both the baseline hazards a_j and $b_j, j = 1, 2, 3$ and the whole path of F between time t_0 and $t_0 + y_2$ (see the discussions in Section 4.3.2).

ii) **Stochastic factor.** Because of the stochastic nature of the longevity, we would also like to model the common factor (F_t) as an unobserved stochastic process, often called dynamic frailty since Duffie et al. (2009). For the comparison with the deterministic exponential specification above, we will assume in applications that the dynamics of the stochastic factor F is a Cox-Ingersoll-Ross (CIR) process [see Cox et al. (1985)] :

$$dF_t = -mF_t dt + \sigma\sqrt{F_t} dW_t, \tag{III-16}$$

where $\sigma > 0$, $m > 0$, W is a standard Brownian motion, and the initial condition is $F_{\min t_0} = 1$, where $\min t_0 := 0$, say, is the birth date of the first cohort.

This CIR model includes the deterministic model as a limiting case. If $\sigma = 0$, then the solution of the differential equation (III-16) is $F_t = \exp(-mt)$. Thus the CIR model is just introducing uncertainty around the deterministic exponential model. Therefore, this CIR process still has a nonstationary feature, which reflects the longevity phenomenon.

The advantage of introducing a stochastic specification of the factor over a deterministic, say, exponential specification, is that we can quantify the uncertainty of both the model fit

and the future evolution. These uncertainties should be taken into account when pricing LTC insurance contracts, computing the regulatory required capitals and performing stress tests [see the discussion in Keilman et al. (2002) for macropolicy implications].

The choice of a CIR process has several other advantages. Firstly, it guarantees the positivity of the intensity functions $\lambda_1, \lambda_2, \lambda_{2|1}$ when functions $a_j, \tilde{b}_j, j = 1, 2, 3$ are nonnegative. Secondly, it allows for closed form expressions of the log-likelihood function under an appropriate approximation scheme by using the affine property of the process.

Appendix B.5 summarizes the basic properties of this CIR process, including its existence, the potential hitting time at 0 and its behavior afterwards, as well as its discrete time counterpart, which is an autoregressive gamma process (ARG).

III.4.3 Nonparametric identification

Let us now discuss the identification issue. For expository purpose, we consider the Markov specification (III-11).

Deterministic exponential factor

Let us first consider the case where the factor F is deterministic and exponential, and the intensity of X_3 given X_1 is Markov. Assume that for each cohort, at the age origin $y_2 = 0$, the proportion of people already in LTC is null, and $F_{t_0} = 1$ for some pre-specified value of t_0.

Property III.3. Assume that we observe the lifetime of a continuum of cohorts of individuals indexed by t, where t varies in an open set $]t_0 - \epsilon, t_0 + \epsilon[$ for $\epsilon > 0$, that the six functions $a_j, b_j, j = 1, 2, 3$ are continuous and positive. Then, the parameter m is identified, and we have the following identification results for the six functions :

1. If $b_1 + b_2 = b_3$ for all y, then b_3 can be globally nonparametrically identified ; the others cannot be identified.
2. If there exists constants c, c' such that $b_1 + b_2 - b_3 \geq c > 0$, and $|b_2 - b_3| > c'$ for each age y, then $b_1 + b_2$ is globally nonparametrically identified ; the other functions are at least locally identified.
3. If there exists a constant d such that $b_1 + b_2 - b_3 \leq -d < 0$ for all y, then functions b_3, a_3 are globally identified ; the other functions are at least locally identified.

Proof :See Appendix B.7. □

In other words, the repeated measurement across different cohorts of the nonlinear effect of the longevity factor on the aggregated lifetime behavior allows for identifying both the functional parameters and the longevity factor. The assumption that at origin, the proportion of people already in LTC is null is an implicit condition of our model and is already used in Equation (III-7). The assumption that all the functions are continuous means that, the entry into LTC is the only possible mortality jump during one's lifetime. The observation of a continuous-valued covariate t is also a standard assumption in the identification literature of survival models [see e.g. Abbring and van den Berg (2003a)] and of treatment effects [see Abbring and van den Berg (2003b) Proposition 2.3.4]. Indeed the proof of identification of m relies on the same "identification at zero" argument as in these papers. Nevertheless our identification result is not a consequence of theirs. Indeed this literature assumes that the time of treatment is observable and usually consider the mixed proportional hazard (MPH) specifications. For longevity models, the specification of the intensities cannot be multiplicative in the observable regressor, due to the need of a limiting model for the far future [see e.g. system (III-15)].

Stochastic factor

Let us now consider the identification of the Markov model with a stochastic factor. Loosely speaking a (functional) parameter is identifiable if it can be consistently estimated. Thus the notion of identification depends on the assumed asymptotics. For our problem, this is a double asymptotics, in which both the number T_0 of observed generations and the number of individuals observed in each generation tend to infinity. In the limiting case of this double asymptotics, the family of survivor functions $S_2(y, t_0)$ given in (III-13) is asymptotically known, that is, we can reconstitute the set of survivor functions given the existing factor path [14]. To summarize we have the following Proposition.

Property III.4. It is equivalent to consider the identification of the intensity components $a_1, b_1, ...$ in a model with stochastic factor, or to consider the identification problem for a model with (unconstrained) deterministic factor, where the factor path coincides with the realized path.

Let us now consider system (III-10). This is a system of equations indexed by y_2 and t_0, which has to be solved w.r.t. functions $a_1, b_1, a_2, b_2, a_3, b_3, F_t$. This system is in general over-identified,

14. For an asymptotics in T_0, with one observed individual in each cohort, say, it would only be possible to reconstitute the integrated survivor function $\overline{S}_2(y, t_0) := \mathbb{E}[S_2(y, t_0)]$, where the expectation is taken with respect to the stochastic future factor path.

except for some special factor paths such as deterministic exponential path. But since (F_t) is a diffusion process, the probability of reduced rank is zero. Thus we have the following Proposition :

Property III.5. Functions $a_1, b_1, a_2, b_2, a_3, b_3$ are locally identifiable, *a.s.*, that is except for a negligible set of factor paths.

The analysis of identification with unobserved stochastic dynamic frailty is completely different from the analysis in standard treatment effect models. Indeed, in models with treatment effects, the unobserved heterogeneity is individual and represented by a scalar or vector random variable. In our framework the longevity factor is a process, therefore much more complex. Nevertheless, the cross-sectional asymptotics allows for eliminating the uncertainty on this factor, that is for replacing the process by its underlying trajectory (Proposition 4). Then the observation of a large number of cohorts introduce the orthogonal dimensions leading to identification (Proposition 5).

Finally, wherever $a_j, b_j, j = 1, 2, 3$ are identifiable, from granularity theory [see e.g. Gagliardini and Gouriéroux (2014)], we can also identify the realized factor path, and then the parameters of the factor dynamics.

III.5 Applications

Under the restrictions introduced in Section 4.2, the scalar and functional parameters of the joint model for longevity and LTC are in general identifiable from lifetime data only. However the lifetimes are also partially observed due to censoring phenomena. In this section we consider the different specifications for models with deterministic or stochastic factors, and derive the likelihood functions, when the entry into LTC is unobserved and the lifetime is right censored.

In our model, the intensity function of the observed variable Y_2 depends in a non Markovian way on all the past of factor F. But under the specifications of the factor that we consider, the likelihood function admits closed form formula when an appropriate discretization scheme is used. We also approximate the functionals $a_j, b_j, j = 1, 2, 3$ by parametric splines. We denote by θ the set of all parameters including both the splines parameters and the parameters characterizing the factor dynamics.

III.5.1 The likelihood function

Model with deterministic factor

Let us first consider the basic model with a deterministic factor $F_t = e^{-mt}$. We denote by $i, i = 1, ..., n$, the individuals and assume that the set of latent variables $(X_{1,i}, X_{2,i}, X_{3,i})$, $i = 1, ..., n$ are independent with identical joint distribution, which depends on the generation only. Then the individual lifetimes $Y_{2,i}, i = 1, ..., n$ are also independent with a distribution depending on t_0 only. Taking into account the right censoring of the lifetimes, the log-likelihood function is :

$$\log l(Y_2, \theta) = \sum_{t_0} \Big\{ \sum_{i \in \mathcal{I}^u_{t_0}} \log f_2(y_{2,i}, t_0, \theta) + \sum_{i \in \mathcal{I}^c_{t_0}} \log S_2(y_{2,i}, t_0, \theta) \Big\}, \tag{III-17}$$

where $\mathcal{I}^u_{t_0}$ (respectively $\mathcal{I}^c_{t_0}$) is the set of uncensored (resp. censored) individuals in generation t_0, $y_{2,i}$ denotes either the observed failure time if the individual is not censored, the censoring time, otherwise, and θ denotes the parameter.

Model with dynamic frailty

The expression of the log-likelihood is similar as (III-17), except that the terms f_2, S_2 should be integrated with respect to the path of factor (F_t). More precisely, we define $\overline{S}_2(y_{2,i}, t_0, \theta) = \mathbb{E}[S_2(y_{2,i}, t_0, F)]$ the integrated survivor function, where $S_2(y_{2,i}, t_0, F)$ is the survivor function conditional on the path of the factor (F_t) and with expression given by (III-13). Similarly we define $\overline{f}_2(y_{2,i}, t_0, \theta) = \mathbb{E}[f_2(y_{2,i}, t_0, F)]$. Then we get :

$$\log l(Y_2, \theta) = \sum_{t_0} \Big\{ \sum_{i \in \mathcal{I}^u_{t_0}} \log \overline{f}_2(y_{2,i}, t_0, \theta) + \sum_{i \in \mathcal{I}^c_{t_0}} \log \overline{S}_2(y_{2,i}, t_0, \theta) \Big\}. \tag{III-18}$$

This expression can be theoretically calculated in continuous time, but at the cost of numerically solving ordinary Riccati differential equations [15]. A simpler way is to approximate the continuous time model with its time-discretized version. This is useful when the available data are collected in discrete time, which is actually the case. More precisely, assume that the intensity functions are constant [16] between two neighboring integer dates : for all x and the integer part of x, $n = \lfloor x \rfloor$,

15. This treatment is standard in the literature of term structure of interest rates and credit spreads with affine underlying factors, see e.g. Duffie et al. (2000).

16. This necessitates also to replace the continuous time CIR process with its time-discretized version, which is an ARG process. See Appendix 5.

say, we have :

$$\lambda_1(x) = \lambda_1(n), \qquad \lambda_2(x) = \lambda_2(n), \qquad \lambda_{2|1}(x) = \lambda_{2|1}(n).$$

Then we get the link between the intensities in continuous and discrete time :

$$\mathbb{P}[X_1 > n+1 \mid X_1 > n] = 1 - \exp(-\lambda_1(n)),$$

and similarly for the other duration variables. The log-likelihood function is therefore approximately :

$$\log l(Y_2, \theta) = \sum_{t_0} \Big\{ \sum_{i \in \mathcal{I}^u_{t_0}} \log f_2^{\text{disc}}(y_{2,i}, t_0, \theta) + \sum_{i \in \mathcal{I}^c_{t_0}} \log S_2^{\text{disc}}(y_{2,i}, t_0, \theta) \Big\}, \tag{III-19}$$

where f_2^{disc} and S_2^{disc} are discrete time approximations of the p.d.f. and the survival function, respectively. They are calculated by first writing the corresponding p.d.f. and survival function $f_2^{\text{disc}}(y_{2,i}, t_0, \theta, F)$ and $S_2^{\text{disc}}(y_{2,i}, t_0, \theta, F)$ conditional on factor path F. Then the dynamic frailty F is integrated out :

$$f_2^{\text{disc}}(y_{2,i}, t_0, \theta) = \mathbb{E}[f_2^{\text{disc}}(y_{2,i}, t_0, \theta, F)], \qquad S_2^{\text{disc}}(y_{2,i}, t_0, \theta) = \mathbb{E}[S_2^{\text{disc}}(y_{2,i}, t_0, \theta, F)].$$

We give in Appendix B.3.2 the expressions of these expectations. They can be written in terms of the Laplace transform of process F, and have closed form for affine processes such as the CIR process (otherwise, the calculation of the log-likelihood requires simulation of the factor paths and is numerically cumbersome).

III.5.2 The data

The methodology of the previous subsections is now applied to a set of observations from the Human Mortality Database (HMD). The HMD was created to provide detailed mortality and population data to researchers, students, policy makers, and others, interested in the history of human longevity. It is maintained by the University of California, Berkeley, and the Max Planck Institute for Demographic Research in Rostock, Germany (see the official website `http://www.mortality.org`).

For instance, for France, the database gives, for each gender and each cohort t_0 since 1737,

the size of the Population-at-Risk and the number of deaths [17] at each integer age, from 0 to $\min(2009 - t_0, 110)$. We use data from age 50 until age 110, and for cohorts starting from 1900. For the oldest cohort (1900), our period of observation begins in 1950 to avoid the period of World War II, and finishes in 2010 ; for the youngest cohort (1958), the observation begins in 2009 and finishes in 2010, which creates the right censoring effect.

Let us now provide summary statistics of the French male population. Because of the longevity phenomenon, the distribution of lifetime is shifting to higher ages. This can be illustrated by the increase of cross-sectional life expectancy [18]. Because of the right censoring, the computation of the real, cohort-based longitudinal life expectancy involves the choice of a predictive model (and will be calculated in Section III.6), while the cross-sectional quantities are model-free, but they do not measure the real expected duration for any cohort. Nevertheless they are still widely used for simplicity. We plot in Figure III-2 the mean age at death observed in a same calendar year.

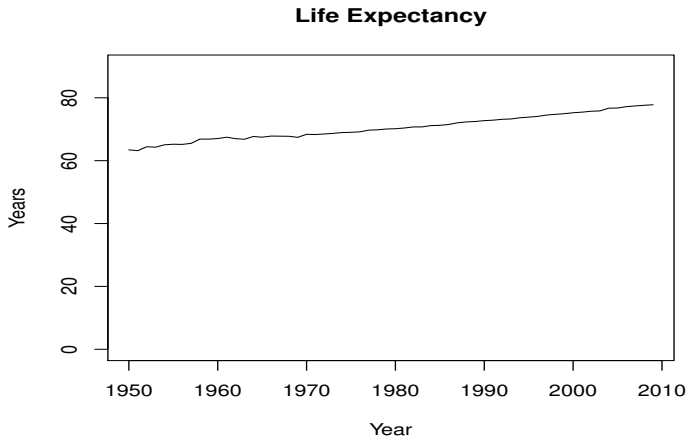


FIGURE III-2: Evolution of the life expectancy at birth for deaths occurring in the same year.

During the past 40 years, the cross-sectional life expectancy for French males has been steadily rising at a rate of approximately 0.25 years, that is 3 month per year. For year 2011, the cross-sectional life expectancy is around 78 years for male, which is about 6 years lower than that of French females', and the latter is also rising at a similar pace.

The longevity phenomenon results in a significant increase of the proportion of seniors in the population, which will potentially need LTC. Figure III-3 shows, for each year, the dependency ratio, that is, the ratio between the size of the old people population (aged 65 or above) and that

17. As a consequence, the corresponding estimates of the mortality intensity function are available as well.
18. Also called period life expectancy in demography.

of the productive population (aged between 15 and 64). This statistics is widely used to measure the pressure on the productive population.

FIGURE III-3: The dependency ratio by year.

The dependency ratio has consistently increased during the last three decades. This is expected to continue as the Baby Boomers reach their retirement ages. This phenomenon spells a huge threat to the sustainability of the social security system and of the pension funds.

III.5.3 Markov model with deterministic exponential factor

We estimate the model introduced in Section III.5.1 on the French male data. We consider the population of males who survive up to age 50. As we suppose an homogeneous population [19], the left censoring is easily taken into account in the log-likelihood function by changing the date origin, which is now 50 instead of age 0.

The model is completed by approximating the functions $a_j, b_j, j = 1, 2, 3$ by linear splines :

Assumption 1. Markov model

i) The function $a_1(x_1)$ is a linear spline for $x_1 \in]50, 110[$ with two knots at 60 and 70 and is null on the interval $]50, 60]$.

ii) The function $b_1(x_1)$ is such that $\tilde{b_1}(x_1) = b_1(x_1)\exp(-mx_1)$ is a linear spline on $]50, 110]$ with two knots at 60 and 70 and is null on the interval $]50, 60]$.

iii) The function $a_2(x_2)$ is a linear spline for $x_2 \in]50, 110[$ with two knots at 80 and 90.

19. By homogeneous population we mean a population without multiplicative unobserved heterogeneity as in Vaupel et al. (1979). Since we assume that at the beginning of the observation ($y = 50$) nobody is in LTC, there is no heterogeneity linked to the initial autonomy status neither.

iv) The function $b_2(x_2)$ is such that $\tilde{b}_2(x_2) = b_2(x_2)\exp(-mx_2)$ is a linear spline on $]50, 110[$ with two knots at 80 and 90.

v) The function $a_3(x_3|x_1) = a_3(x_3 + x_1)$ is a linear function of the current age $x_3 + x_1$, for $x_3 + x_1 \in]60, 110[$.

vi) The function b_3 is such that $\tilde{b}_3(x_3|x_1) = b_3(x_3|x_1)e^{-m(x_3+x_1)}$ is a linear function of $x_3 + x_1$ function for $x_3 + x_1 \in]60, 110[$.

Let us now comment on these assumptions. We specify the baseline hazards under the age-period decomposition [see equation (III-11)]. The linear spline specification is a nonparametric method to approximate the baseline functions. It would be possible to choose more knots, but numerical experiments show that this offers little benefit and may induce over-parameterization and less robust results. Empirically we find that other parametric specifications, such as exponential splines, can also fit the model relatively well. We show in Appendix B.3 that the linear spline specification provides closed form expressions of the log-likelihood function in some special cases. Assumptions *v*) and *vi*) written on the transition intensity function λ_3 are Markov conditions.

Let us now discuss the choice of the age range used in our estimation. We only look at people who survive age 50, since the mortality pattern at younger ages is significantly different from that of higher ages. In general, there are very few people in LTC before age 60 ; therefore we assume that functions a_1 and b_1 are null between 50 and 60. Our model is written up to age 110, which is approximately the current limit age of the human being [20]. It would equally be possible to restrict the observation window to, say, ages 50-90 : this would (very slightly) improve the fit of the model, but will prevent us from predicting the residual life expectancy.

The following Lexis diagram illustrates the relationship between the cohort, age and calendar years. The observed part of the history of each cohort is represented by a full 45° line whose left and right boundaries are respectively the age of the beginning and end of the observation (due to either right censoring). As for the censored parts, they are plotted in thick dashed lines. Of all the cohorts, we distinguish two cases :

— Cohorts born before 1900 (for instance cohort 1870 in the plot) are not taken into account in the estimation. Indeed, their post age 50 history is impacted by the second world war, the aftermath of which marks a strong regime switch in terms of mortality improvement.

— Cohorts after 1900 are right censored, and the censoring age equals $\min(110, 2010 - t_0)$

20. The oldest living human is currently a 116 years old man.

for a cohort born in t_0. For instance, for cohort 1930, only the data from age 50 to 80 are used.

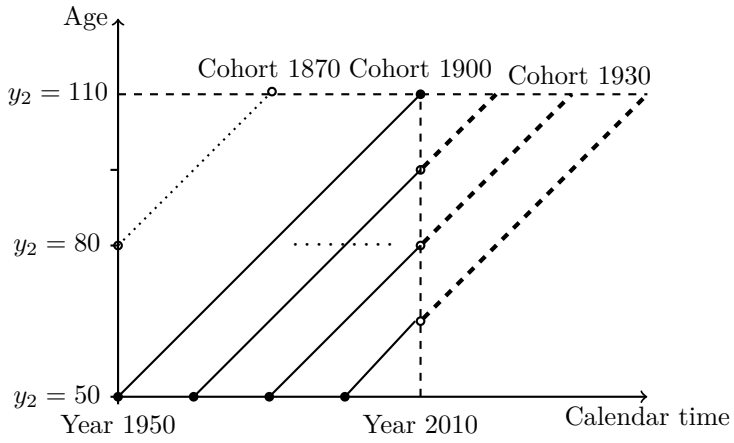


FIGURE III-4: Lexis diagram of cohorts and their observability. The study period ranges from year 1950 to 2010.

The following table gives a summary of the linear splines $a_1, \tilde{b}_1, a_2, \tilde{b}_2, a_3, b_3$ in terms of their value at origin as well as their slopes between different knots.

TABLE III.1: Parameters of the linear spline functions

	value at 50	slope between 50, 60	slope between 60, 70	slope between 70, 80	slope between 80, 90	slope between 90, 110
$a_1(x)$	0	0	w_1	w_2	w_2	w_2
$a_2(x)$	w_3	w_4	w_4	w_4	w_5	w_6
$b_1(x)$	0	0	w_7	w_8	w_8	w_8
$\tilde{b}_2(x)$	w_9	w_{10}	w_{10}	w_{10}	w_{11}	w_{12}
$a_3(x)$	w_{13}	w_{14}	w_{14}	w_{14}	w_{14}	w_{14}
$\tilde{b}_3(x)$	w_{15}	w_{16}	w_{16}	w_{16}	w_{16}	w_{16}

Under Assumption 1, the set of all parameters is $\theta = (w_1, w_2, ..., w_{16}, m)$. For brevity the value of the estimator, the goodness of fit, as well as the discussion of this model are given in Appendix B.4.1. We first compute the model implied intensity function of Y_2 and compare it to the historical data. Besides, we can also plot the evolution of the latent hazard functions, as well as the implied evolution of the proportion of people in long term care (i.e. prevalence), that is, the distribution of the unobserved heterogeneity.

III.5.4 Semi-Markov model with deterministic exponential factor

In the previous Markov model, we have assumed that the mortality intensity for a person in LTC depends only on its current age. A more realistic and intuitive assumption is that it depends also on the age of entry into LTC z, or equivalently, on the time elapsed since this entry $x - z$. Therefore, in this section, we consider the following semi-Markov assumption :

Assumption 2. Semi-Markov model

i) Functions $a_1(x)$, $b_1(x)$, $a_2(x)$ and $b_2(x)$ are specified in the same way as in Assumption 1.

ii) Function $a_3(x - z|z)$ and $b_3(x - z|z)\exp(-mx)$ are linear both in x and z :

$$\begin{cases} a_3(x-z|z) & = c_{0,a} + c_{1,a}(x-z) + \beta_1(z-60), \\ b_3(x-z|z)\exp(-mx) & = c_{0,b} + c_{1,b}(x-z) + \beta_2(z-60). \end{cases}$$

The additional parameters β_1, β_2 characterize the non Markovian feature. For this semi-Markov model, the set of parameters becomes :

$$\theta = (w_1, w_2, ..., w_{12}, c_{0,a}, c_{1,a}, c_{0,b}, c_{1,b}, \beta_1, \beta_2, m).$$

The estimation and discussion are gathered in Appendix B.4.2.

III.5.5 Model with dynamic frailty

Let us finally replace, in the previous semi-Markov model, the deterministic dynamic factor by a (common) dynamic frailty, as explained in Subsection III.5.1. The parameters of the model, including those of the CIR process [equation (III-16)], m, σ, and those of the baseline hazard functions a_j, b_j, $j = 1, 2, 3$, are estimated jointly by maximizing the log-likelihood function given by equation (III-19). Since the model with deterministic factor is the limiting case of the model with dynamic frailty, we can choose the initial value of the numerical algorithm used to optimize the likelihood function as $w = (w^*, 0)$, where w^* is the value of the maximum likelihood estimator of the semi-Markov model with deterministic factor derived in Section III.5.4. We report in Table III.2 the value of the estimator w.

TABLE III.2: Estimator of the model with dynamic frailty, all parameters are significant at 1% level.

variable	estimator
w_1	0.000693 (***)
w_2	0.002568 (***)
w_3	0.005693 (***)
w_4	0.000168 (***)
w_5	0.003672 (***)
w_6	0.018114 (***)
w_7	0.000425 (***)
w_8	0.002639 (***)
w_9	0.002827 (***)
w_{10}	0.001485 (***)
w_{11}	0.002958 (***)
w_{12}	0.023078 (***)
$c_{0,a}$	0.177399 (***)
$c_{0,b}$	0.009781 (***)
$c_{1,a}$	0.003288 (***)
$c_{1,b}$	0.005822 (***)
β_1	0.004991 (***)
β_2	0.004737 (***)
σ	0.020561 (***)
m	0.034579 (***)

To look at the goodness of fit, we compute the intensity function of the lifetime variable Y_2 for each cohort, when the dynamic frailty is integrated out. More precisely, we first compute the survivor function of the lifetime at different times by integrating out the whole history of the dynamic frailty, and then we calculate the hazard function by computing its minus log-derivative :

$$h(y_2) = \lim_{h\to 0} \frac{\mathbb{P}[y_2 \leq Y_2 < y_2 + h]}{h} = -\frac{\partial}{\partial y_2} \log \mathbb{E}\Big[S_2(y_2|\theta, F)\Big] = \frac{\mathbb{E}\Big[f_2(y_2|\theta, F)\Big]}{\mathbb{E}\Big[S_2(y_2|\theta, F)\Big]}. \qquad \text{(III-20)}$$

We display in Figure III-5 the intensity function of Y_2 and compare its values to the observed values from the data.

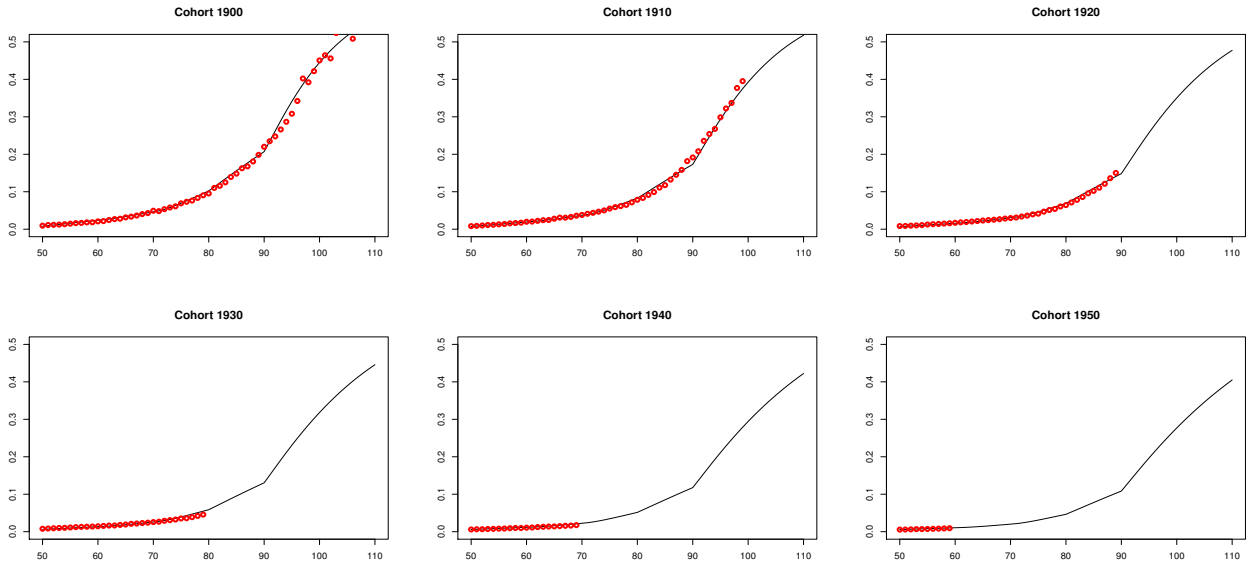


FIGURE III-5: Hazard function of the lifetime variable. Dotted line : historical data. Full line : the model (for both the past and future years).

Once the parameters are estimated, we infer the path of unobserved frailty process (F_t). This is useful for several reasons. First, after filtering out the unobserved frailty process, we can check the specification of its dynamics (CIR process), as well as the goodness of fit of the model in terms of observable mortality rates. Second, its values can be used for predicting the future mortality and the LTC transition probability, which depend on the frailty process.

There are at least two ways to filter out this unobserved process. First, the observed mortality rates can be written as (nonlinear) functions of the values of the unknown frailty and of parameters. We may invert these equations to obtain the values of the frailty process after replacing the parameter by its maximum likelihood estimate. This methodology is widely used in Finance, [see e.g. Chen and Scott (1993)]. However, since functions $\overline{f}_2(y_2, t_0, \theta), \overline{S}_2(y_2, t_0, \theta)$ depend on the frailty path in a non Markovian and nonlinear way, and the number of unknown frailty values is quite large when the process covers the period 1951-2009, this approach is numerically cumbersome. For the same reason, nonlinear filtering methods [see e.g. Gagliardini et al. (2012)] are equally forbidden.

The second method is based on simulations of the factor path after substituting the estimated parameters to their true values. More precisely, we simulate a certain number of paths of the frailty process conditionally on both the estimated value of the parameter and on the observations

$Y_{2,i}, i \in \mathcal{I}^u \cup \mathcal{I}^c$, that are either the dates of death or the right censoring ages of all individuals. This is done by Gibbs sampling, as in Duffie et al. (2009). Appendix B.6 gives the details of this methodology. In Figure III-6, we plot, for each year, the simulated factor mean $\mathbb{E}[F_t|\theta, Y_2]$ conditional on all the observed $Y_{2,i}, i \in \mathcal{I}^u \cup \mathcal{I}^c$. For comparison, we also plot the deterministic path $\mathbb{E}[F_t|\theta] = e^{-m(t-1950)}$, where m is the trend parameter of the CIR process.

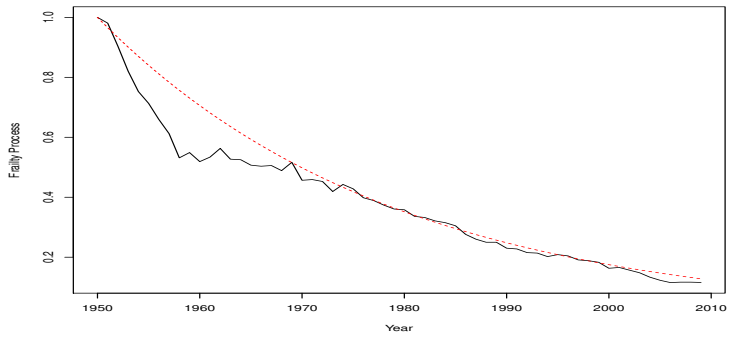


FIGURE III-6: Simulated factor mean (full line) and the deterministic path (dotted line).

As expected, the path features a nonstationary (decreasing) trend, which corresponds to the longevity phenomenon. The filtered factor mean is different from the deterministic path, that is $\mathbb{E}[F_t|\theta, Y_2] \neq \mathbb{E}[F_t|\theta]$, because of the conditioning on the information Y_2. Indeed for most dates t, we observe empirically that $\mathbb{E}[F_t|\theta, Y_2] < \mathbb{E}[F_t|\theta]$. This result was expected, since the longevity phenomenon favors paths of the CIR process that feature a more pronounced decrease. The filtered paths of the factor can also be used to calculate the conditional intensity of Y_2, that is $\lambda_2(y_2|\theta, F)$, where the values of factor F are replaced by their filtered values. Not surprisingly, for each of its simulated paths, we get very satisfactory fit to the observed lifetime intensity similarly as in Figure III-5. These figures are omitted due to lack of space.

This factor does not have the same influence on the different latent intensities $\lambda_1(x_1, t_0)$, $\lambda_2(x_2, t_0)$, $\lambda_{2|1}(x_3, t_0|x_1)$; indeed these effects depend on the ratios $a_1(x_1)/b_1(x_1)$, $a_2(x_2)/b_2(x_2)$, $a_3(x_3|x_1)/b_3(x_3|x_1)$, who depend themselves on the values of x_1, x_2, x_3. These values can be used to compare the improvement speed of different intensity functions. This was also true for the two previous models with deterministic factor. For instance, for the Markov model with deterministic factor, we see from Figure B-3 that the reduction of $\lambda_{2|1}$ at age $x_3 + x_1 = 100$ is less important (about 50 %) than that of λ_2 (about 67 %).

III.5.6 Comparison of the models with deterministic and stochastic factors

The three models, that are the Markov and semi-Markov model with deterministic factor as well as the semi-Markov model with stochastic factor all provide satisfactory fits. The maximized log-likelihoods are respectively : -38710240, -38709452, -38704065, and the corresponding values of the BIC are : 77420764, 77419205 and 77408448. It was expected that the semi-Markov model with deterministic factor (resp. the semi-Markov model with stochastic factor) has a higher likelihood than the nested semi-Markov model with deterministic factor (resp. Markov model with deterministic factor), but the difference is rather small. However, the comparison between the semi-Markov models with deterministic and stochastic factor requires more care. Indeed the standard BIC criterion is not necessarily the appropriate measure to compare the performance of the two models in terms of risk prediction and risk management. For instance we have already mentioned that a model with deterministic common factor will likely underestimate the risk. The next section offers a further comparison of these models in terms of prediction.

III.6 Prediction of individual LTC and mortality risks

Once the model is estimated from the lifetime data, we can infer for each individual the value of the unobserved variables given the observed ones. We consider below an individual of cohort t_0 at calendar date $t_0 + y_2$. For a model with deterministic factor, it is rather easy to deduce the expressions of the predictive distributions ; for a model with dynamic frailty, some expectations, such as the hazard function of the lifetime variable (see Equation (III-20)), admit explicit forms after integrating out the frailty process, but confidence intervals have to be computed by simulation. More precisely, for each simulated past history of process F obtained from the Gibbs sampler (see Subsection III.5.5), we simulate its future path and obtain the predictive distributions conditional on the whole factor path, whose formulas are similar as for the model with deterministic factor. This procedure is repeated to obtain the prediction intervals. The prediction problem depends on the observed variables. We have the following situations :

i) If the individual is already dead, we know the value of Y_2, but have to predict the potential date of entry into LTC Y_1 as well as the latent variables X_1, X_2, X_3.

ii) If the individual is still alive and we have no information on his/her autonomy state, except

that $Y_2 > y_2$, we have to predict Y_1, X_1, X_2, X_3 and Y_2.

iii) If the individual is autonomous, that is, $X_1 > y_2, X_2 > y_2$, we have to predict Y_1, Y_2, X_1, X_2, X_3,

and so on. We first derive explicit prediction formulas for a model with deterministic factor. Then we consider the prediction of future risks in Case *iii*) for the French males, by both the Markov model with deterministic factor and the semi-Markov model with dynamic frailty. These quantities are calculated for different cohorts, but for expository purpose we omit the cohort index t_0. Since the individual observations are independent, we can perform the computation independently for each individual. For expository purpose we omit the individual index i.

III.6.1 Case *i*)

Let us first consider the case of predicting unobserved variables, which include the variable Y_1, and the latent variables (X_1, X_2, X_3), conditional on the complete observation of Y_2. The expressions of the predictive distributions are derived below.

Conditional distribution of Y_1 given Y_2. This distribution has a density with respect to the measure $\delta_0 + \lambda_{]0,y_2[}$, where δ_0 is the point mass at 0. This density is :

$$f(Y_1 = 0|Y_2 = y_2) = \frac{f(0, y_2)}{f(0, y_2) + \int_0^{y_2} f(y_1, y_2)dy_1} = \mathbb{P}(Y_1 = 0|Y_2 = y_2), \qquad \text{if } Y_1 = 0,$$

and

$$f(Y_1 = y_1|Y_2 = y_2) = \frac{f(y_1, y_2)}{f(0, y_2) + \int_0^{y_2} f(y_1, y_2)dy_1}, \qquad \text{if } Y_1 \neq 0,$$

where $f(\cdot, \cdot)$ is the joint density function [see equations (III-4) and (III-5)].

Conditional distribution of (X_1, X_2, X_3) given Y_2. This conditional distribution has two components on domain $\mathcal{D}_3 = \{(x_1, x_2, x_3) \in \mathbb{R}_{\geq 0}, x_1 + x_3 = y_2, x_2 \geq y_2\}$, and $\mathcal{D}_4 = \{(x_1, x_2, x_3) \in \mathbb{R}_{\geq 0}, x_2 = y_2, x_1 \geq y_2\}$, respectively. Both domains are subsets of a hyperplane. The joint

distribution admits a density with respect to the measure $\lambda_{\mathcal{D}_3} + \lambda_{\mathcal{D}_4}$. This density is :

$$g(x_1, x_2, x_3 | Y_2 = y_2) = \frac{g(x_1, x_2, y_2 - x_1)}{f_2(y_2)}, \qquad \text{on domain } \mathcal{D}_3,$$

and

$$g(x_1, x_2, x_3 | Y_2 = y_2) = \frac{g(x_1, y_2, x_3)}{f_2(y_2)}, \qquad \text{on domain } \mathcal{D}_4.$$

III.6.2 Case *ii)*

Let us now consider the case when only the information $Y_2 > y_2$ is available.

Conditional distribution of Y_1 given $Y_2 > y_2$. This conditional distribution has three components corresponding to three different cases : $Y_1 = 0$, $Y_1 < y_2$ and $Y_1 > y_2$. It has a density with respect to the measure $\delta_0 + \lambda_{]0,y_2[}$, and this density is :

$$f(y_1 | Y_2 > y_2) = \frac{\lambda_1(t) e^{-\Lambda_1(t) - \Lambda_2(t) - \Lambda_{2|1}(y_2 - t|t)}}{\displaystyle\int_0^{y_2} \lambda_1(t) e^{-\Lambda_1(t) - \Lambda_2(t) - \Lambda_{2|1}(y_2 - t|t)} dt + e^{-\Lambda_1(y_2) - \Lambda_2(y_2)}}, \qquad \text{on domain } \{y_1 \in]0, y_2]\},$$

$$f(y_1 | Y_2 > y_2) = \frac{\lambda_1(t) e^{-\Lambda_1(t) - \Lambda_2(t)}}{\displaystyle\int_0^{y_2} \lambda_1(t) e^{-\Lambda_1(t) - \Lambda_2(t) - \Lambda_{2|1}(y_2 - t|t)} dt + e^{-\Lambda_1(y_2) - \Lambda_2(y_2)}}, \qquad \text{on domain } \{y_1 \in]y_2, \infty[\},$$

and

$$f(0 | Y_2 > y_2) = \frac{\displaystyle\int_{y_2}^{\infty} \lambda_2(t) e^{-\Lambda_1(t) - \Lambda_2(t)} dt}{\displaystyle\int_0^{y_2} \lambda_1(t) e^{-\Lambda_1(t) - \Lambda_2(t) - \Lambda_{2|1}(y_2 - t|t)} dt + e^{-\Lambda_1(y_2) - \Lambda_2(y_2)}}, \qquad \text{if } Y_1 = 0.$$

It is easily checked that this function $f(\cdot | Y_2 > y_2)$ sums up to 1 and we have :

$$\int_0^{y_2} f(y_1 | Y_2 > y_2) dy_1 = p(y_2),$$

that is the prevalence at age y_2 [see Equation (B-7)].

Conditional distribution of Y_2 given $Y_2 > y_2$. This is already characterized by the hazard function of Y_2 (see e.g. Equation (B-6) for the Markov model).

The conditional distribution of (X_1, X_2, X_3) given $Y_2 > y_2$ can be obtained similarly and its expression is omitted.

III.6.3 Case *iii)*

Let us now assume that the available information set is $X_1 > y, X_2 > y$. A special case is when $y = 50$, since any individual enrolled in the study at this age is autonomous [21], and we are interested in the prediction of Y_1 and Y_2. First, let us compute the probability that a person will enter the LTC during his or her lifetime, given autonomy up to age y. For each cohort, this probability is given by :

$$\mathbb{P}(Y_1 > 0 | X_1 > y, X_2 > y) = \frac{\int_y^\infty \lambda_1(x) e^{-\Lambda_1(x)-\Lambda_2(x)} dx}{e^{-\Lambda_1(y)-\Lambda_2(y)}}. \tag{III-21}$$

This probability is called the cumulative incidence (at age $Y_2 = \infty$).

Other interesting quantities include the residual life expectancy with (potential) LTC.

$$\begin{aligned} e_1(y) &= \mathbb{E}[Y_2 - y | X_1 > y, X_2 > y] \\ &= \frac{\int_y^\infty (x_2 - y)\lambda_2(x_2) e^{-\Lambda_1(x_2)-\Lambda_2(x_2)} dx_2}{e^{-\Lambda_1(y)-\Lambda_2(y)}} \\ &\quad + \frac{\int_y^\infty \Big(x_1 + \int_0^\infty x_3 \lambda_{2|1}(x_3|x_1) e^{-\Lambda_{2|1}(x_3|x_1)} dx_3 - y\Big) \lambda_1(x_1) e^{-\Lambda_1(x_1)-\Lambda_2(x_1)} dx_1}{e^{-\Lambda_1(y)-\Lambda_2(y)}}, \end{aligned}$$

as well as the residual life expectancy without LTC (or Healthy Life Years [22]) defined by :

$$e_2(y) = \mathbb{E}[\min(X_1, X_2) - y | X_1 > y, X_2 > y] = \frac{\int_y^\infty (x - y)\Big(\lambda_1(x) + \lambda_2(x)\Big) e^{-\Lambda_1(x)-\Lambda_2(x)} dx}{e^{-\Lambda_1(y)-\Lambda_2(y)}}.$$

This term is very popular among sociologists. Indeed, the issue of increasing life expectancy in good health has become a huge concern for policy makers in recent years in developed countries.

Then we can compute the difference of these two terms, which is the expected duration spent in the potential LTC state [23]. It is of particular interest to insurance companies or public social

21. Since the transition intensity into LTC is null before age 60.
22. This term is introduced by Eurostat, the statistical service of the European Commission. It is calculated in a cross-sectional way while our $e_1(y), e_2(y)$ are longitudinal measures. An alternative terminology is the Disability-Free Life Expectancy (DFLE) [see e.g. Imai and Soneji (2007)].
23. For a person who never entered LTC during its lifetime, this duration is zero.

security plans, since it impacts the expected cost of an LTC insurance policy in a direct way. We have :

$$\begin{aligned} & e_1(y) - e_2(y) = \mathbb{E}[X_3 \mathbb{1}_{Y_1>0} | X_1 > y, X_2 > y] \\ &= \frac{\int_y^\infty \Big(\int_0^\infty x_3 \lambda_{2|1}(x_3|x_1) e^{-\Lambda_{2|1}(x_3|x_1)} dx_3 \Big) \lambda_1(x_1) e^{-\Lambda_1(x_1)-\Lambda_2(x_1)} dx_1}{e^{-\Lambda_1(y)-\Lambda_2(y)}}. \end{aligned} \tag{III-22}$$

In general, the term $\int_0^\infty x_3 \lambda_{2|1}(x_3|x_1) e^{-\Lambda_{2|1}(x_3|x_1)} dx_3$, that is, the expected residual lifetime upon entry at age x_1, depends on x_1 and cannot be factored out.

Let us now calculate the three quantities above for different values of age y and cohort t_0. For expository purpose, we use the Markov model with deterministic factor and the semi-Markov model with dynamic frailty. For the latter one, 90% confidence bounds are also provided, that are, the 5% and 95% quantiles of the variable $\mathbb{P}[X_1 < X_2 | X_1 > y, X_2 > y, F]$, which is calculated for each simulated factor path F. Figure III-7 displays the evolution of the probability of entering into LTC during its lifetime given survival up to age 50 as a function of the cohort t_0. The value of y is set to 50 years.

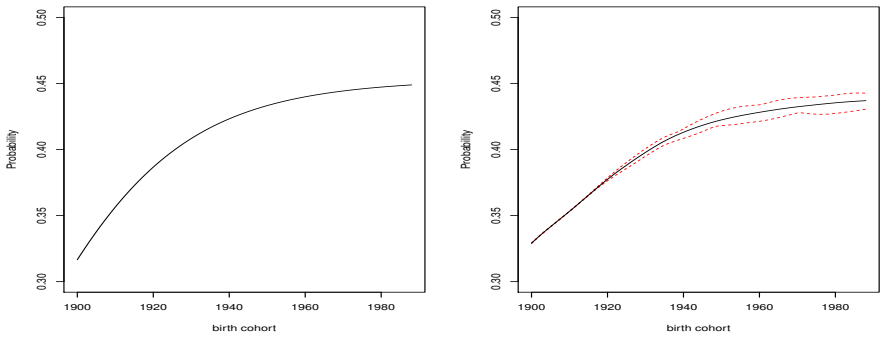


FIGURE III-7: Evolution of the probability of entering into LTC during its lifetime as a function of the cohort. Left panel : the Markov model with deterministic factor, right panel : the semi-Markov model with dynamic frailty ; full line : the expected value, that is when frailty is integrated out, dashed lines : the 90% confidence bounds.

The Markov model predicts a slightly higher probability of entering into LTC than the semi-Markov model with dynamic frailty, but in both cases, this probability is increasing in cohort. For instance, the latter predicts that this probability is around 0.33 for the oldest cohort (born in 1900) and will be around 0.43 for the cohort 1980. Theses probabilities are in line with the projection based on LTC use history of a sample of Americans by Spillman and Lubitz (2002),

who predict that in 2020, the probability of a 65-year-old [24] ever entering a nursing home to will increase to 46 %. The result is also to be compared to Figure B-7 in Appendix, where we plot the proportion of people in LTC at any ages, which is decreasing in cohort [25]. For the semi-Markov model with dynamic frailty, the uncertainty, measured by the bandwidth of the confidence interval, is increasing in cohort : for the cohort 1900, the bandwidth is very close to (but not strictly equal to) zero, and becomes quite large for, say, cohort 1980. Indeed, the variation of the filtered past path is considerably smaller than the variation of its predicted future path because of the conditioning with respect to the information of Y_2. For cohort 1900, its history depends only on the filtered past history of the factor F, whereas for cohort 1980 it depends also on the future evolution of the path.

Let us now plot in the same figures the evolution of the residual life expectancies (with and without LTC) for an individual in good health at age 50, for cohorts born from 1900 to 1988.

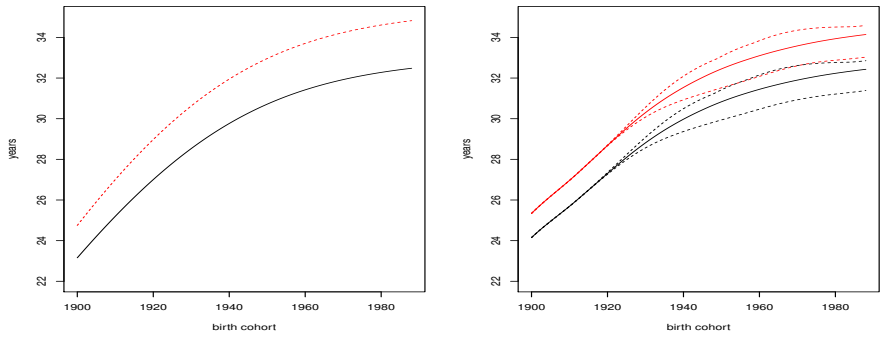


FIGURE III-8: Evolution in t_0 of the residual life expectancy, with potential LTC (dashed line) and without (full line) LTC, at age 50. Left panel : the Markov model, right panel : the semi-Markov model with dynamic frailty ; full lines : the expected values, dashed lines : the 90% confidence bounds ; the three upper curves are for the life expectancy with potential LTC.

For a French male aged 50 in 2010, the residual life expectancy with potential LTC is around 33 years with the semi-Markov model. The curve of the residual life expectancy with potential LTC is slightly concave, and increases with an average improvement rate of around 0.1 year per annum. The difference between the two curves, which directly impacts the expected cost of an LTC insurance contract, is (slowly) increasing.

24. Which is roughly of the same order than the probability for a 50-year-old given the relatively lower intensities between age 50 and 65.
25. Similarly, the probability of surviving until a given age, either with or without disability, is increasing.

Finally, let us calculate the uncertainty of the following quantities for a finite population :

$$\frac{1}{n}\sum_{i=1}^{n} Y_{2,i,t_0}, \qquad \frac{1}{n}\sum_{i=1}^{n} \min(X_{1,i,t_0}, X_{2,i,t_0}), \tag{III-23}$$

where Y_{2,i,t_0} [resp. $\min(X_{1,i,t_0}, X_{2,i,t_0})$] is the future death age (resp. age of either losing autonomy or dying directly) for the individual i aged 50 in, say, year $\tau = 2010$. In other terms, these two sums correspond to the average residual lifetime with (resp. without) LTC for a homogeneous portfolio of n individuals. We are interested in calculating their Value-at-Risk $VaR(\alpha)$, where $\alpha \in]0, 1[$.

The computation of these VaR can be done by simulation, but this is very time consuming when the size of the portfolio is large. Nevertheless, it can be approximated by using the granularity theory [see e.g. Gagliardini and Gouriéroux (2014)]. For the model with deterministic factor factor, the distribution of the quantities in (III-23) are approximately Gaussian by the Central Limit Theorem. For the model with dynamic frailty, conditional on each simulated factor path, these quantities are still approximately Gaussian ; therefore their unconditional distribution is approximately a mixture of, say, M Gaussian distributions, where M is the number of simulated factor paths. When the size of the portfolio goes to infinity, the asymptotic VaR, i.e. cross-sectional asymptotic (CSA) VAR, provides the undiversifiable component of the risk. This CSA VaR is easily calculated : for the model with deterministic factor, it is equal to zero ; for the model with dynamic frailty, it equals the 95% quantile of the conditional expectation $e_1(y|F) = \mathbb{E}[Y_2|X_1 > y, X_2 > y, F]$ (resp. $e_2(y|F) = \mathbb{E}[\min(X_1, X_2)|X_1 > y, X_2 > y, F]$). These quantities have already been calculated (see Figure III-8).

To illustrate this approach, let us take $n = 10, 100, \infty$, and $\alpha = 0.05, 0.95$. The confidence bounds are displayed in Table III.3.

TABLE III.3: 90% confidence bounds for the average residual lifetime for a portfolio of n individuals who are 50 years old in 2010.

Empirical mean of Y_2	$n = 10$	$n = 100$	$n = \infty$
Markov model without frailty	33.12, 33.60	33.29, 33.44	33.36 ± 0
Semi-Markov model with frailty	31.95, 33.86	32.03, 33.85	32.18, 33.78
Empirical mean of $\min(X_1, X_2)$	$n = 10$	$n = 100$	$n = \infty$
Markov model without frailty	30.98, 31.47	31.15,31.30	31.22 ± 0
Semi-Markov model with frailty	30.45, 32.10	30.47, 32.16	30.59, 32.08

For both empirical means, the confidence interval is larger for the model with (common) frailty, which incorporates the uncertainty of the frailty process (both its future and past), whereas the Markov model without frailty assumes it equal to zero. The model with frailty is therefore more reliable from the insurer point of view.

III.6.4 Comparison with real data on LTC

Let us finally compare the model-based prediction with data on LTC from a large insurance company. Such private proprietary database usually concern the customers and are not representative of the whole population. They are subject to selection biases due to both the behavior of the company and of the customers. Let us discuss the expected bias for the analysis of LTC.

— Since the LTC insurance market is young and small, products are not very differentiated. Thus the insurance company will try, for a given price of the contract, to select the least risky customers [26]. Thus we expect that in this database, the time spent in LTC is smaller than for the whole population.

— On the other hand, the standard economic literature insists on the role of adverse selection which tends to increase the average risk profile of the customers. However, this standard argument seems to be not valid in the LTC framework, a finding also confirmed by Finkelstein and McGarry (2006). They attribute this to the offsetting effect of selection into the market and find evidence that wealthier individuals and individuals who exhibit more cautious behavior are both more likely to have LTC insurance coverage and less likely to use LTC. Indeed, in insurance problems with irreplacable objects, individuals' utility function is in general state-dependent [see e.g. Dionne (1982), Karni (1983)], i.e. with a higher risk aversion in the LTC state. The preference to be better covered in this state will imply an increased demand. On the other hand, the weak effect of the adverse selection could also be partially explained by the long-term nature of the risk, which makes it difficult for individuals to exploit asymmetric information. In the same direction will be the bias coming from the income effect since the customers who can afford a private insurance are likely to have a higher income than the national average.

To summarize, we expect that the endogenous selection by both the insurance company and the customers are going in the same direction of overweighting of the best risks, i.e. smaller

26. For instance, many insurance companies believe that living with one's partner, as well as being smoker, are indicators of small time spent in LTC.

probability of entering into in LTC in the database w.r.t. the whole population.

The database concerns a specific insurance product with only one LTC state ; it has been launched in 1994 and sales continued up to [27] 2000, but the database is maintained even after that date. There are about 15000 male policyholders [28], the majority of whom were born between 1925 and 1940 (see Figure III-9 for a histogram of the cohort of all policyholders) and bought the contract in their 60's. Thus they are quite young at the end of the observation period, that is 2014. As a consequence, observations are heavily right censored. Indeed, 20 % individuals died without LTC and only 5 % entered into LTC before the end of the observation period, with a potentially further censored final death date ; the other observations are completely censored. No events are observed beyond age 90. The portfolio size is not sufficient to conduct a real cohort-specific analysis and the individuals from different cohorts are aggregated.

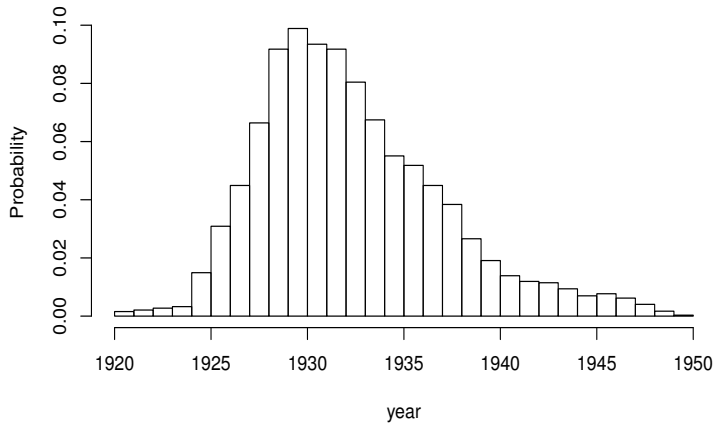


FIGURE III-9: Histogram of birth cohort of all policyholders.

27. After 2000, the company launched a new product with significant changes of policy terms ; therefore the new product cannot be compared directly to the original one.

28. The size of the portfolio is rather reduced with respect to the French population. Nevertheless, it is believed to be one of the largest and most reliable databases from one of the largest reinsurance companies in the world. This illustrates the difficulties of the insurance industry in providing comparable LTC products, and in maintaining quality databases.

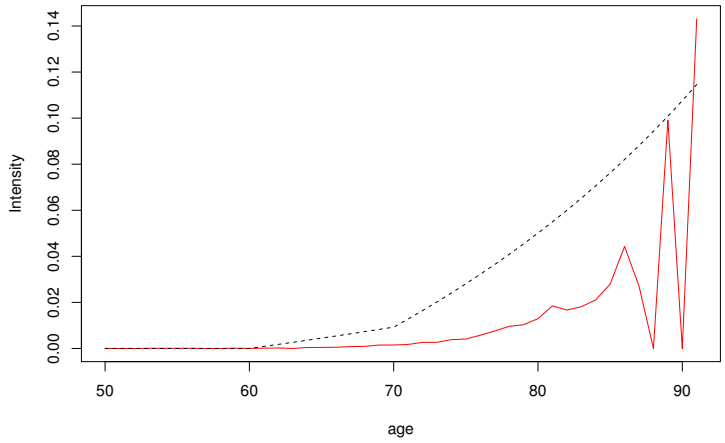


FIGURE III-10: Comparison between the intensity of entry into LTC implied by the model (for general population) and that observed on the insurance data. Dashed line : the model ; full line : the data.

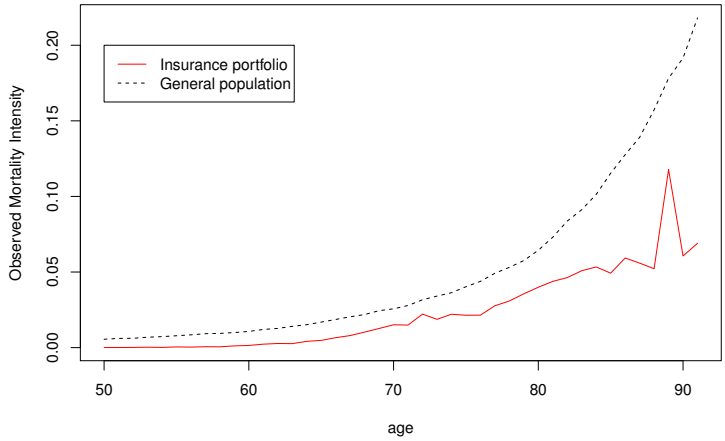


FIGURE III-11: Comparison between the observed mortality intensity of the two populations.

Figure III-10 compares the intensities of entering into LTC computed for the set of policyholders and deduced from the estimated model for the general population. For the insurance portfolio, the estimated intensity is $\hat{\lambda}_1(x) = -\frac{d}{dx}\log \hat{S}_1(x)$, where $\hat{S}_1(x)$ is the Kaplan-Meier estimator of the marginal survivor function of the entry into LTC. For the model based intensity at each age x, we took a weighted sum of $\Big(\lambda_1(x|t_0)\Big)$ for different cohorts t_0, where the weights are determined by the share of each cohort among individuals that survive up to age x. This allows

us to correct the longevity bias of the aggregated portfolio. Figure III-10 shows that our model predicts a slightly higher intensity of entry into LTC for general population than that observed from the insurance data, especially for lower ages. This difference can be partly explained by the endogenous selection of policyholders by the insurance company and the choice of individuals to buy such a contract. Whereas the entry in LTC is exogenous, the enrollment in a private LTC coverage is endogenous (see the discussion at the beginning of the current subsection). To further confirm the selection effects, Figure III-11 plots the aggregated mortality intensity (without distinguishing the autonomy state) for both the policyholders and the general population. The huge discrepancy between the two curves suggests that the insured population has a much better health than the general population, and, therefore are likely to have a lower intensity of entry into LTC [29]. This comparison shows the difficulty in taking into account the available LTC insurance data, when estimating the models, due to the poor data quality and the endogenous selection.

III.7 Conclusion

In this paper we proposed a new methodology to predict the probabilities of entering into LTC along with the mortality intensities with or without LTC using solely the lifetime data. In this modeling, the entry into LTC is characterized by a jump in the mortality intensity. In some sense we get a model based implied LTC state which can be used as long as the data on LTC are either unavailable, or weakly reliable, or under endogenous selectivity. This implicit state may differ from that of a specific LTC database [30] and it would be interesting to compare the hypothetical date of entry in LTC with the different dates of losing Eating, Dressing, ... abilities, when longitudinal data will become available and reliable. This may lead to change the definition of the Instrumental Activities of Daily Living, as well as the design of LTC insurance products.

Our model is based on minimal [31] observability and thus assumes a single LTC state. In some cases it may be attempting to include other observed information, such as the regular measurement of various individual health indicators, or even direct observation of the LTC use. In the latter case, we will “force" the latent state to match certain characteristics of an observable LTC state. The inclusion of such information is theoretically possible, but since it often comes

29. In other words we assume a positive correlation between LTC and mortality risks. See the beginning of this subsection or Murtaugh et al. (2001) for a discussion of this assumption.
30. Which is logical, especially given the lack of a universal definition and the poor quality of existing databases.
31. Although repeated across different cohorts.

from a different database for a smaller population sample and/or a shorter period, its effective use requires additional, case-dependent care. This can be an area of further research when appropriate database becomes available.

Finally, the joint statistical analysis of entry into LTC and mortality is a requested step, before checking if individual LTC risk is really insurable by insurance companies, or if it is profitable to combine mortality and LTC risks into a joint insurance product [32].

Appendix B.1 Expressions of the survivor function and the p.d.f. of the lifetime variable Y_2

The expression of the p.d.f. of Y_2 is obtained by integrating out the joint density with respect to y_1. We get :

$$\begin{aligned} f_2(y_2) &= \int f_2(y_1, y_2) dy_1 \mathbb{1}_{0<y_1<y_2} + f(0, y_2) \\ &= \int_0^{y_2} \lambda_1(t)\lambda_{2|1}(y_2 - t|t) e^{-\Lambda_1(t)-\Lambda_2(t)-\Lambda_{2|1}(y_2-t|t)} dt + \lambda_2(y_2) e^{-\Lambda_1(y_2)-\Lambda_2(y_2)}. \end{aligned}$$

Let us now check the expression of the survivor function by computing its derivative. We get :

$$\begin{aligned} -\frac{dS_2(y_2)}{dy_2} &= -\lambda_1(y_2) e^{-\Lambda_1(y_2)-\Lambda_2(y_2)} \\ &\quad + \int_0^{y_2} \lambda_1(t)\lambda(y_2 - t|t) e^{-\Lambda_1(t)-\Lambda_2(t)-\Lambda(y_2-t|t)} dt \\ &\quad + \Big[\lambda_1(y_2) + \lambda_2(y_2)\Big] e^{-\Lambda_1(y_2)-\Lambda_2(y_2)} \\ &= f_2(y_2). \end{aligned}$$

Appendix B.2 Technical lemmas

Lemma B.1. Given $a, b, \alpha, \beta > 0$, let us consider the function g defined by :

$$g(y) = a \exp(-\alpha y) - b \exp(-\beta y), \qquad y \in]0, \infty[;$$

32. For instance, Murtaugh et al. (2001) argue that based on the assumption that the two risks are positively correlated, then combining the two risks would significantly lower the overall insurance premium, increase the attractiveness of the products, and thus also limit the adverse selection.

then g is a survivor function if and only if $a = b + 1$ and $\frac{b}{b+1}\beta < \alpha < \beta$.

Proof : The necessary and sufficient condition for g to be a survivor function is $g(0) = 1$ and g is decreasing. The first condition gives $a = b + 1$. Let us now focus on the second condition. The derivative of g is :

$$\frac{d}{dy}g(y) = -\alpha a \exp(-\alpha y) + b\beta \exp(-\beta y).$$

Therefore g is a survivor function if and only if :

$$a = b + 1 \text{ and } \frac{a\alpha}{b\beta} \geq \exp((\alpha - \beta)y), \qquad \forall y > 0,$$

or equivalently $a = b + 1$ and $\frac{b}{b+1}\beta < \alpha < \beta$. □

Lemma B.2. Given $a, b > 0$, let us consider the function g defined by :

$$g(y) = (1 + by)e^{-ay}, \qquad y \in]0, \infty[;$$

then g is a survivor function if and only if $a \geq b$.

Proof : The condition $g(0) = 1$ is satisfied. Therefore g is a survivor function if and only if :

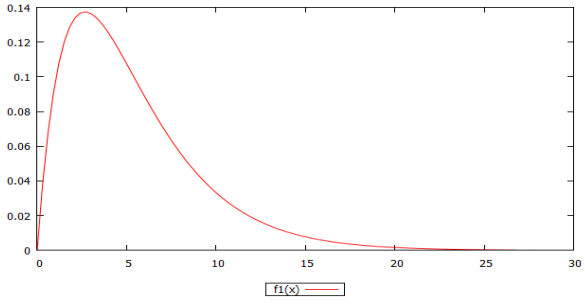
$$\frac{dg}{dy} = -e^{-ay}(aby + a - b) \geq 0, \qquad \forall y > 0,$$

or equivalently $a \geq b$. □

As an illustration, we plot below the corresponding p.d.f. of the survivor function :

$$S(y) := \frac{\lambda_1 + \lambda_2}{\lambda_1 + \lambda_2 - \lambda_{2|1}} e^{-\lambda_{2|1} y} - \frac{\lambda_{2|1}}{\lambda_1 + \lambda_2 - \lambda_{2|1}} e^{-(\lambda_1 + \lambda_2) y},$$

where we set the parameters as following : $\lambda_1 = 0.1, \lambda_2 = 0.3, \lambda_{2|1} = 0.35$.



Appendix B.3 Expression of the log-likelihood function

B.3.1 Model with deterministic factor

In this section we give the detailed expression of the log-likelihood function (III-17) in the model with deterministic factor. For expository purpose let us start by considering the Markov model. The semi-Markov case is slightly more complicated but is based on the same principle. By using the age-cohort decomposition, we have,

$$\begin{aligned}
&f_2(y_{2,i}, t_0, \theta) \\
&= \Big(a_3(y_{2,i}) + \tilde{b}_3(y_{2,i})F_{t_0}\Big) \int_0^{y_{2,i}} [a_1(x) + \tilde{b}_1(x)F_{t_0}] \exp\Big(- \int_0^x [a_1(s) + \tilde{b}_1(s)F_{t_0}]ds \\
&\qquad - \int_0^x [a_2(s) + \tilde{b}_2(s)F_{t_0}]ds - \int_x^{y_{2,i}} [a_3(s) + \tilde{b}_3(s)F_{t_0}]ds\Big)dx \\
&\quad + \Big(a_2(y_{2,i}) + \tilde{b}_2(y_{2,i})F_{t_0}\Big) \exp\Big(- \int_0^{y_{2,i}} [a_1(x) + \tilde{b}_1(x)F_{t_0}]dx - \int_0^{y_{2,i}} [a_2(x) + \tilde{b}_2(x)F_{t_0}]dx\Big),
\end{aligned} \tag{B-1}$$

and

$$\begin{aligned}
S_2(y_{2,i}, t_0, \theta) = &\int_0^{y_{2,i}} [a_1(x) + \tilde{b}_1(x)F_{t_0}] \exp\Big(- \int_0^x [a_1(s) + \tilde{b}_1(s)F_{t_0}]ds \\
&\qquad - \int_0^x [a_2(s) + \tilde{b}_2(s)F_{t_0}]ds - \int_x^{y_{2,i}} [a_3(s) + \tilde{b}_3(s)F_{t_0}]ds\Big)dx \\
&\quad + \exp\Big(- \int_0^{y_{2,i}} [a_1(x) + \tilde{b}_1(x)F_{t_0}]dx - \int_0^{y_{2,i}} [a_2(x) + \tilde{b}_2(x)F_{t_0}]dx\Big),
\end{aligned} \tag{B-2}$$

where we have changed the time origin ($t = 0$ corresponds to age 50) to account for the left censoring.

Let us now derive the closed form expression of these functions under the linear spline Assumption 1. For any integer value of $y_{2,i}$, consider the interval $[y_{2,i} - 1, y_{2,i}]$. On this interval, functions $a_j, \tilde{b}_j, j = 1, 2, 3$ are all linear in x and the factor $F_{t_0} = e^{-mt_0}$ does not depend on x, we can write $a_1(x)+\tilde{b}_1(x)F_{t_0} = s_1x+i_1$, $a_2(x)+\tilde{b}_2(x)F_{t_0} = s_2x+i_2$, and $a_3(x)+\tilde{b}_3(x)F_{t_0} = s_3x+i_3$, where $s_1, s_2, s_1, i_1, i_2, i_3$ are constants and can be expressed by the coefficients of the linear splines

and of $F_{t_0} = \exp(-mt_0)$. Let us now write :

$$S_2(y_{2,i}, t_0, \theta) = e^{-\int_0^{y_{2,i}} [a_3(s)+\tilde{b}_3(s)F_{t_0}]ds} \int_0^{y_{2,i}} [a_1(x) + \tilde{b}_1(x)F_{t_0}] \exp\Big(- \int_0^t [a_1(s) + \tilde{b}_1(s)F_{t_0}]ds$$
$$- \int_0^t [a_2(s) + \tilde{b}_2(s)F_{t_0}]ds + \int_0^t a_3(s) + \tilde{b}_3(s)F_{t_0}ds\Big)dx$$
$$+ \exp(-s_3 y_{2,i}^2/2 - i_3 y_{2,i}), \tag{B-3}$$

where we factored the term $e^{-\int_0^{y_{2,i}} [a_3(s)+\tilde{b}_3(s)F_{t_0}]ds}$ out of the first integral so that the integrand of the remaining integral does not depend on the upper bound $y_{2,i}$. This new integral can be calculated recursively by using the relationship : $\int_0^{y_{2,i}} = \int_0^{y_{2,i}-1} + \int_{y_{2,i}-1}^{y_{2,i}}$. We get :

$$\int_{y_{2,i}-1}^{y_{2,i}} [a_1(x) + \tilde{b}_1(x)F_{t_0}] \exp\Big(- \int_0^t [a_1(s) + \tilde{b}_1(s)F_{t_0}]ds - \int_0^t [a_2(s) + \tilde{b}_2(s)F_{t_0}]ds + \int_0^t a_3(s) + \tilde{b}_3(s)F_{t_0}ds\Big)dx$$
$$= e^{-s_3 y_{2,i}^2/2 - i_3 y_{2,i}} \int_{y_{2,i}-1}^{y_{2,i}} (s_1 x + i_1) \exp\Big(- (s_1 + s_2 - s_3)(x - y_{2,i} + 1)^2/2 - (i_1 + i_2 - i_3)(x - y_{2,i} + 1)\Big)dx$$
$$+ \exp(-s_3 y_{2,i}^2/2 - i_3 y_{2,i}).$$

The first term is of the form $\int A(x)e^{-B(x)}dx$ with A (respectively B) linear (respectively quadratic). If $s_1 + s_2 - s_3 > 0$, which is often the case, then this term can be expressed in terms of the cumulative distribution function of the normal distribution, therefore $S_2(y_{2,i}, t_0, \theta)$ and $f_2(y_{2,i}, t_0, \theta)$ can be expressed in (quasi) explicit form [33]. For the semi-Markov model, we cannot factor out the term $e^{-\int_0^{y_{2,i}} [a_3(s)+\tilde{b}_3(s|x)F_{t_0}]ds}$ because of the dependence on x. As a consequence the recursive formula is not valid, but for fixed $y_{2,i}$, the integrand of the integral in (B-1) and (B-2) is still of the form $\int A(x)e^{-B(x)}dx$, where A and B are piecewise linear (resp. quadratic) therefore the integral can be calculated in explicit form by dividing the integration interval into several subintervals where A and B are linear (resp. quadratic).

B.3.2 The model with dynamic frailty

Let us adopt the discretization scheme described in Subsection 4.2.2, and replace the continuous time process (F_t) by its time discretized version $(F_{[t]})$. whose values at integer times is an

33. Indeed, the cumulative distribution function has no closed form, but its computation is rather fast using standard softwares.

ARG process (see Appendix B.5).

$$\overline{f}_2^{\text{disc}}(y_{2,i}, t_0, \theta) = \mathbb{P}[Y_{2,i} = y_{2,i}] = \mathbb{E}\Big[\mathbb{E}[Y_{2,i} = y_{2,i} \mid F]\Big]$$
$$= \mathbb{E}\Bigg[\sum_{i=0}^{y_{2,i}-1} \Big[1 - e^{-a_1(i) - b_1(i) F_{t_0+i}}\Big] \Big[1 - e^{-a_3(y_{2,i}|i) - b_3(y_{2,i}|i) F_{t_0+y_{2,i}}}\Big]$$
$$\exp\Big(- \sum_{j=0}^{i-1} [a_1(j) + b_1(j) F_{t_0+j}] - \sum_{j=0}^{i-1} [a_2(j) + b_2(j) F_{t_0+j}] - \sum_{j=i+1}^{y_{2,i}-1} [a_3(j|i) + b_3(j|i) F_{t_0+j}]\Big)\Bigg]$$
$$+ \mathbb{E}\Bigg[\Big(1 - e^{-a_2(y_{2,i}) - b_2(y_{2,i}) F_{t_0+y_{2,i}}}\Big) \exp\Big(- \sum_{i=0}^{y_{2,i}-1} [a_1(i) + b_1(i) F_{t_0+i}] - \sum_{i=0}^{y_{2,i}-1} [a_2(i) + b_2(i) F_{t_0+i}]\Big)\Bigg], \tag{B-4}$$

and

$$\overline{S}_2^{\text{disc}}(y_{2,i}, t_0, \theta) = \mathbb{P}[Y_{2,i} > y_{2,i}] = \mathbb{E}\Big[\mathbb{E}[Y_{2,i} > y_{2,i} \mid F]\Big]$$
$$= \mathbb{E}\Bigg[\sum_{i=0}^{y_{2,i}} \Big[1 - e^{-a_1(i) - b_1(i) F_{t_0+i}}\Big] \exp\Big(- \sum_{j=0}^{i-1} [a_1(j) + b_1(j) F_{t_0+j}] - \sum_{j=0}^{i-1} [a_2(j) + b_2(j) F_{t_0+j}]$$
$$- \sum_{j=i+1}^{y_{2,i}} [a_3(j|i) + b_3(j|i) F_{t_0+j}]\Big)\Bigg]$$
$$+ \mathbb{E}\Bigg[\exp\Big(- \sum_{i=0}^{y_{2,i}} [a_1(i) + b_1(i) F_{t_0+i}] - \sum_{i=0}^{y_{2,i}} [a_2(i) + b_2(i) F_{t_0+i}]\Big)\Bigg]. \tag{B-5}$$

These terms are lagged Laplace transform of the process (F_t) and can be calculated in explicit form by iterating the equation :

$$\mathbb{E}[e^{-uF_{t+1}} | \underline{F_t}] = \exp\Big(- \frac{e^{-m} u}{1 + cu} F_t\Big),$$

where $c = \frac{1 - e^{-m}}{2m} \sigma^2$ and u is a nonnegative argument. Again, as for the model with deterministic factor, the computation is faster for the Markov model than for the semi-Markov model, since in the first case, we can factor out the term $\exp(-\sum_{j=0}^{y_{2,i}} [a_3(j|i) + b_3(j|i) F_{t_0+j}])$, which does not depend on i and both $f_2(y_2)$ and $S_2(y_2)$ can be calculated recursively.

Appendix B.4 Estimation results

B.4.1 Markov model with deterministic exponential factor

The model is estimated by maximum likelihood using the R package *DEoptim*. We report below the value of the maximum likelihood estimator, and derive the standard deviation of its components by calculating numerically the inverse of the Fisher Information matrix.

TABLE III.4: Estimation of the Markov model with deterministic exponential factor. All parameters are significant at 1% level.

variable	estimator	standard deviation	t-statistics
w_1	0.000398	0.0000158	25.1 ***
w_2	0.001441	0.0000338	42.7 ***
w_3	0.006955	0.0000256	271.3 ***
w_4	0.00024	0.0000051	47.2 ***
w_5	0.005047	0.0001091	46.3 ***
w_6	0.004713	0.0010629	4.4 ***
w_7	0.000285	0.0000225	12.7 ***
w_8	0.002342	0.0000385	60.8 ***
w_9	0.002037	0.0000408	50 ***
w_{10}	0.000784	0.0000071	110.7 ***
w_{11}	0.00259	0.0001255	20.6 ***
w_{12}	0.015769	0.0010415	15.1 ***
w_{13}	0.228108	0.0166392	13.7 ***
w_{14}	0.242871	0.0192654	12.6 ***
w_{15}	0.005123	0.0007004	7.3 ***
w_{16}	0.004978	0.0006665	7.5 ***
m	0.036432	0.0003179	114.5 ***

With the estimated value of parameter θ, we can derive the estimated intensity function for the lifetime variable Y_2 for a given cohort t_0 and a given age y_2 by using the following formula :

$$\lambda(y_2, t_0, \theta) = f_2(y_2, t_0, \theta)/S_2(y_2, t_0, \theta).$$

This is the mortality intensity, when the unobserved heterogeneity of autonomy status is integrated out. Therefore, it is a weighted average of the intensity functions of the two subgroups : autonomous and non autonomous. Indeed, using the expression of the p.d.f. f_2 and of the survivor

function S_2, we have :

$$\begin{aligned}\lambda(y_2) &= \lambda_{2|1}(y_2)\frac{\displaystyle\int_0^{y_2}\lambda_1(t)e^{-\Lambda_1(t)-\Lambda_2(t)-\Lambda_{2|1}(y_2-t|t)}dt}{\displaystyle\int_0^{y_2}\lambda_1(t)e^{-\Lambda_1(t)-\Lambda_2(t)-\Lambda_{2|1}(y_2-t|t)}dt + e^{-\Lambda_1(y_2)-\Lambda_2(y_2)}} \\ &\quad + \lambda_2(y_2)\frac{e^{-\Lambda_1(y_2)-\Lambda_2(y_2)}}{\displaystyle\int_0^{y_2}\lambda_1(t)e^{-\Lambda_1(t)-\Lambda_2(t)-\Lambda_{2|1}(y_2-t|t)}dt + e^{-\Lambda_1(y_2)-\Lambda_2(y_2)}} \\ &= \lambda_{2|1}(y_2)p(y_2) + \lambda_2(y_2)\Big(1-p(y_2)\Big), \end{aligned} \tag{B-6}$$

where we have omitted the cohort index t_0, as well as the parameter θ to simplify the notations. The weight $p(y_2)$ is the proportion of people in LTC among the whole Population-at-Risk who survive up to a given age y_2 and is given by :

$$\begin{aligned}p(y_2) &= \frac{\displaystyle\int_0^{y_2}\lambda_1(t)e^{-\Lambda_1(t)-\Lambda_2(t)-\Lambda_{2|1}(y_2-t|t)}dt}{\displaystyle\int_0^{y_2}\lambda_1(t)e^{-\Lambda_1(t)-\Lambda_2(t)-\Lambda_{2|1}(y_2-t|t)}dt + e^{-\Lambda_1(y_2)-\Lambda_2(y_2)}} \\ &= \frac{\mathbb{P}[0<Y_1<y_2, y_2<Y_2]}{\mathbb{P}[y_2<Y_2]} \\ &= \mathbb{P}[0<Y_1<y_2|Y_2>y_2]. \end{aligned} \tag{B-7}$$

This probability is the **prevalence** at age y_2 and depends also on the cohort t_0.

Then we can compare the values of this intensity function of Y_2 at each integer age to the historical values of the dataset for the corresponding cohort and age, to look at the goodness of fit of the model in terms of the observed intensity, first by cohort (see Figure B-1), then by age (see Figure B-2). These figures show a rather good fit for the mortality intensities. Then we plot the latent baseline hazard functions λ_1, λ_2, and $\lambda_{2|1}$ (see Figure B-3). The model predicts that the mortality intensity of dependent people is larger than that of autonomous people ($\lambda_{2|1} > \lambda_2$), which is often the case in reality.

We plot also the evolution of the prevalence function $p(y_2, t_0)$ for different cohorts (see Figure B-4).

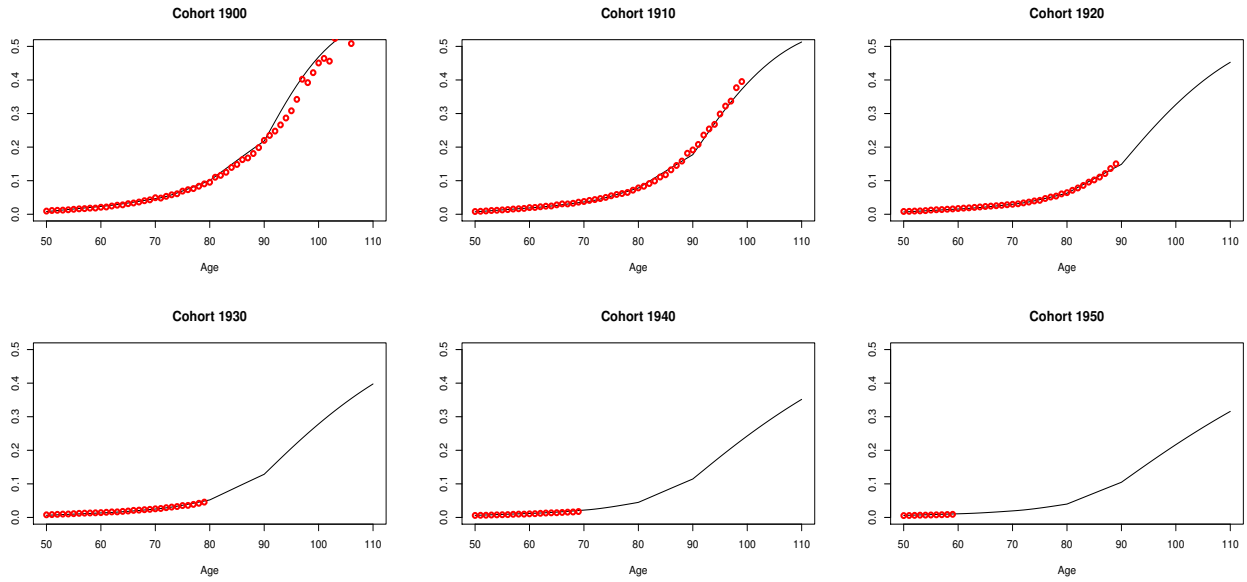


FIGURE B-1: Fit of the observable mortality rates, for six different **cohorts**. Dotted line : historical data. Full line : the model (for both the past and future years). The x coordinate represents the age.

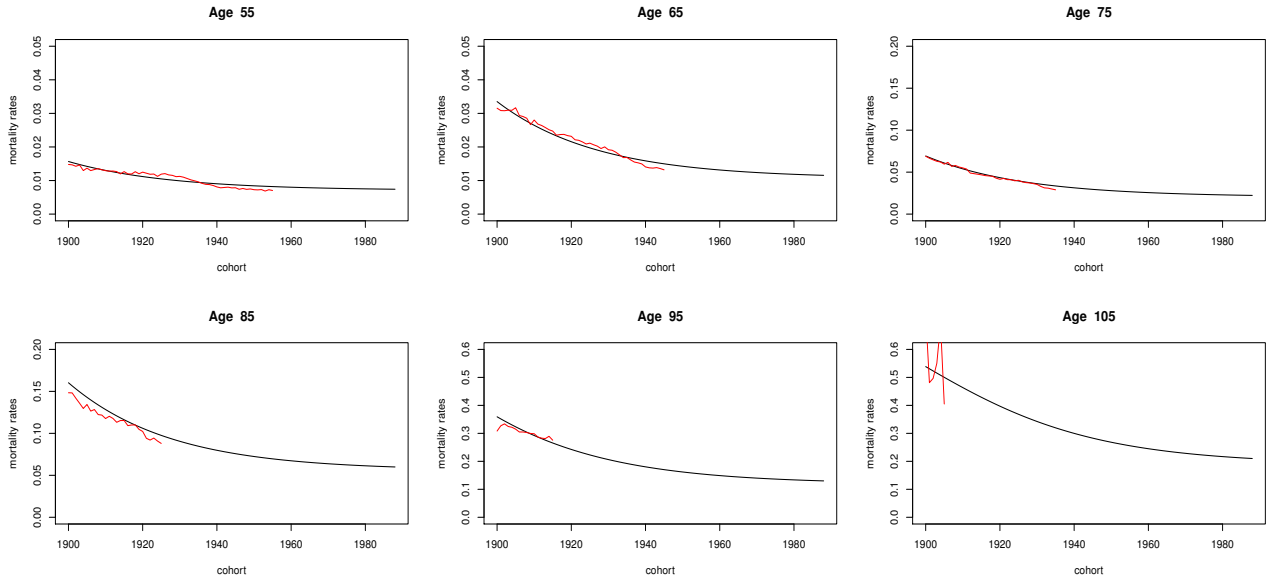


FIGURE B-2: Fit of the observable mortality rates, for nine different **ages**. Dotted line : historical data. Full line : the model (for both the past and future years). The x coordinate represents the cohort.

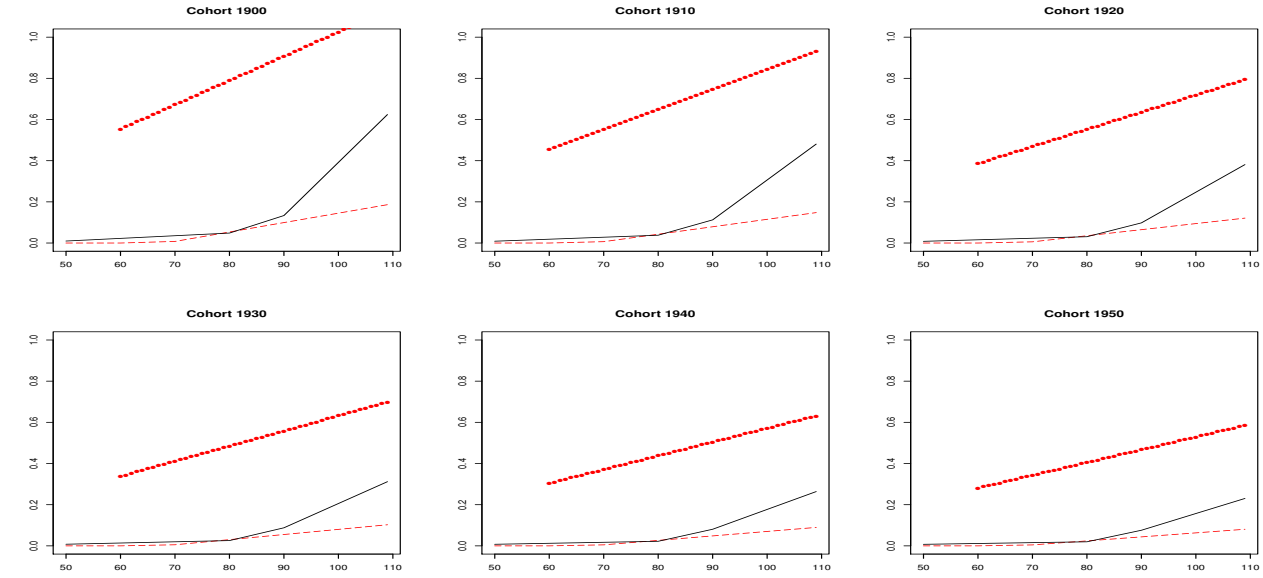


FIGURE B-3: Evolution of the model based baseline hazard functions, respectively $\lambda_1(x)$ (for the intensity of entry, dashed line), $\lambda_2(x)$ (for mortality without LTC, full line) and $\lambda_3(x)$ (mortality of person in LTC, dotted line).

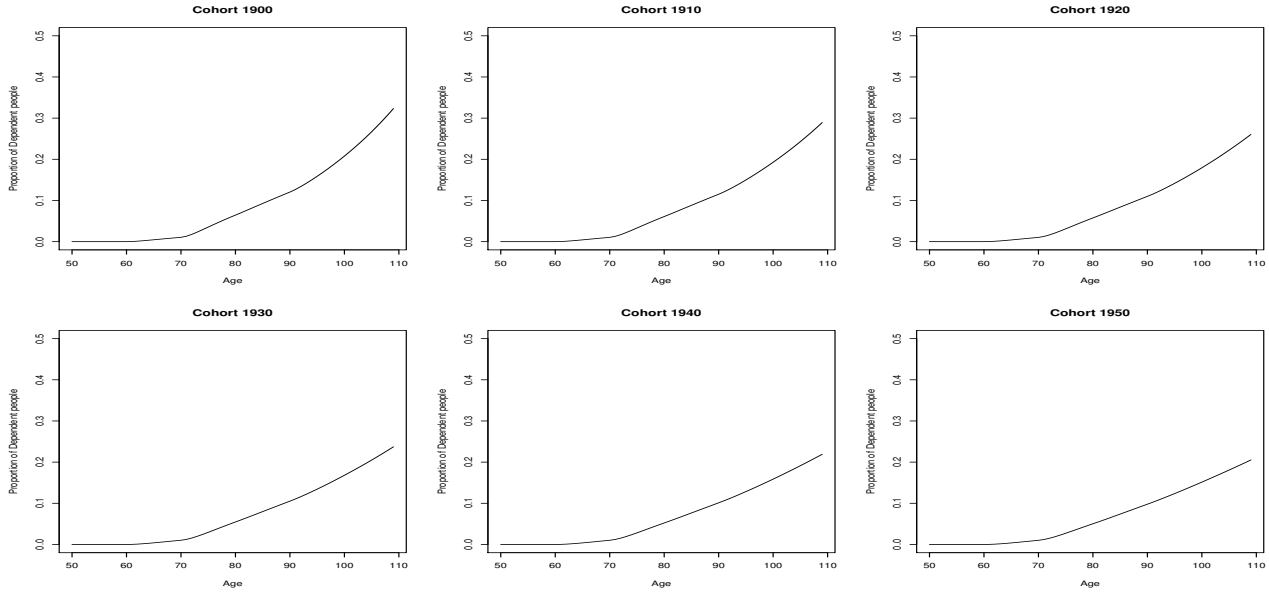


FIGURE B-4: Evolution of the model based proportion of dependent people at a given age for each cohort.

The model predicts that the prevalence begins from 0 at young ages to around 40 percent at age 110 for the cohort 1900, which corresponds roughly to the observed cross-sectional statistics. This prevalence decreases in t_0 for each given age. This proportion reaches 10% at age 82, 85

and 88 for the following cohorts : 1900, 1920, 1940, respectively. This corresponds approximately to an increase of 1.8 months per annum for the age of entry into LTC to be compared with the 3-month increase for the cross-sectional life expectancy.

B.4.2 Semi-Markov model with deterministic exponential factor

As the previous Markov model, the parameter is estimated by maximizing the log-likelihood function. The estimated parameters are reported below :

TABLE III.5: Estimation of the semi-Markov model with deterministic exponential factor ; all parameters are significant at 1% level.

w_1	0.000647 (***)
w_2	0.001983 (***)
w_3	0.005249 (***)
w_4	0.000234 (***)
w_5	0.003322 (***)
w_6	0.014902 (***)
w_7	0.000354 (***)
w_8	0.003278 (***)
w_9	0.002738 (***)
w_{10}	0.001389 (***)
w_{11}	0.003532 (***)
w_{12}	0.020574 (***)
$c_{0,a}$	0.234175 (***)
$c_{0,b}$	0.010442 (***)
$c_{1,a}$	0.0037 (***)
$c_{1,b}$	0.006254 (***)
β_1	0.014494 (***)
β_2	0.020769 (***)
m	0.034201 (***)

To illustrate the fit of the model, we compare for different cohorts the value of the estimated intensity $\lambda(y_2, t_0, \theta)$ with the historical mortality intensity function given by the data (Figure B-5).

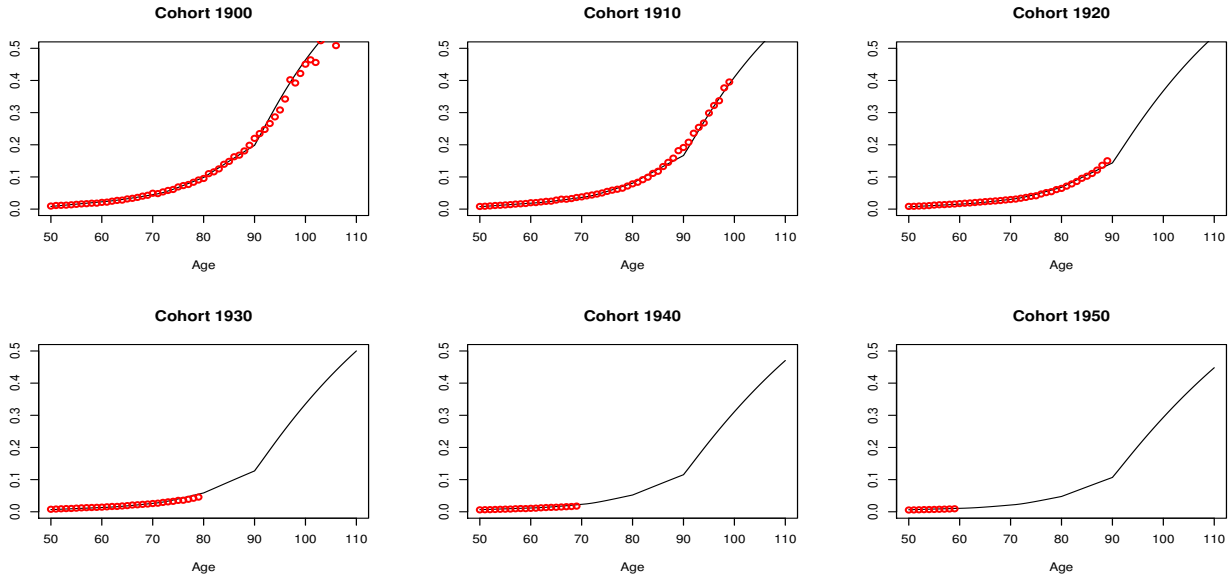


FIGURE B-5: Fit of the observable mortality rates, for six different **cohorts**. Dotted line : historical data. Full line : the model (for both the past and the future years). The x coordinate represents the age.

The semi-Markov model provides also a very good fit. Then we plot (see Figure B-6), for different cohorts, the baseline hazard functions λ_1 and λ_2, since they depend only on the age y_2. For the mortality intensity of people in LTC, we plot, for each cohort, the averaged mortality intensity of all the people aged y_2 in LTC : $\ddot{\lambda}_{2|1}$, say. It is defined for each cohort by :

$$\ddot{\lambda}_{2|1}(y_2) = \frac{\displaystyle\int_0^{y_2} \lambda_1(z)\lambda_{2|1}(y_2 - z|z)e^{-\Lambda_1(z)-\Lambda_2(z)-\Lambda_{2|1}(y_2-z|z)}dz}{\displaystyle\int_0^{y_2} \lambda_1(z)e^{-\Lambda_1(z)-\Lambda_2(z)-\Lambda_{2|1}(y_2-z|z)}dz}.$$

Then we can check that equations (B-6) and (B-7) still hold when we replace $\lambda_{2|1}(y_2)$ by $\ddot{\lambda}_{2|1}(y_2)$. Figure B-7 plots, for several cohorts, the evolution of the proportion of people in LTC.

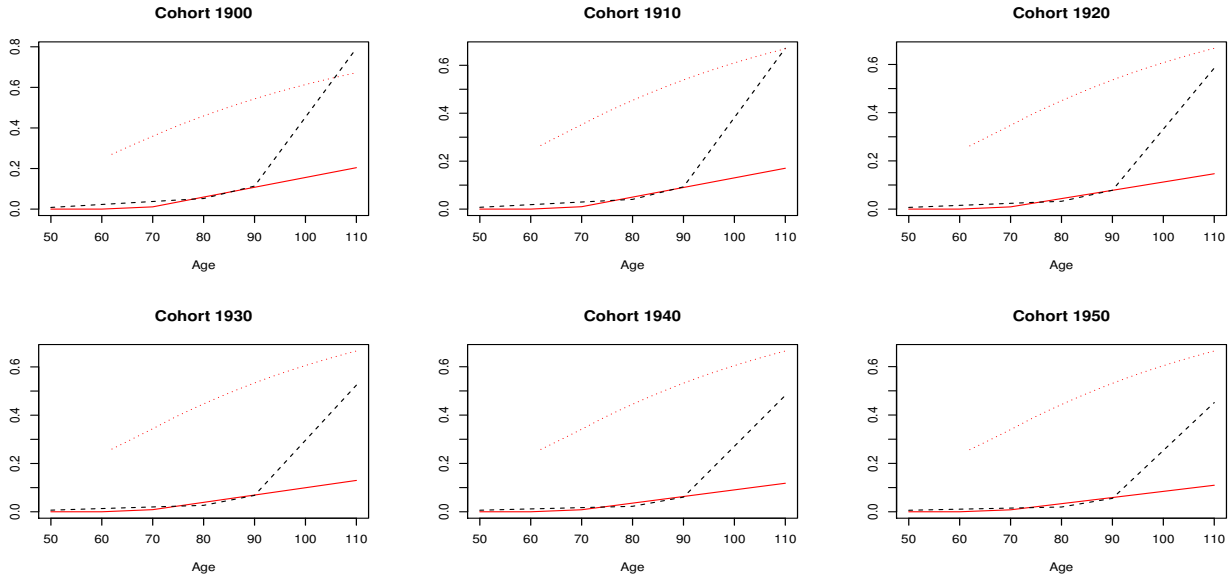


FIGURE B-6: Evolution of the baseline hazard functions, respectively, $\lambda_1(x)$ (for the probability of entering into LTC, dashed line), $\lambda_2(x)$ (for mortality without LTC, full line) and $\ddot{\lambda}_3$ (mortality of people in LTC, dotted line).

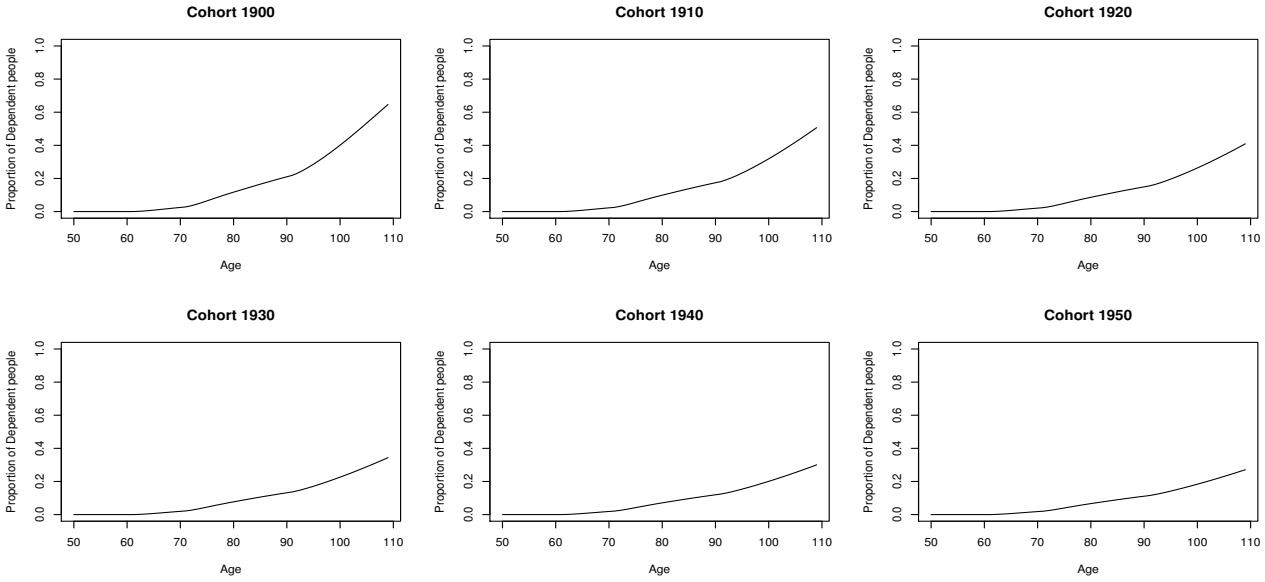


FIGURE B-7: Evolution of model based proportion of people in LTC, for each cohort.

Appendix B.5 Properties of the latent CIR process

This section provides a brief summary of the properties of the CIR process satisfying :

$$dF_t = -mF_t + \sigma\sqrt{F_t}dW_t.$$

Lemma B.3. The stochastic differential equation (SDE) defines a unique strong solution. With probability 1, this solution attains 0 in a stochastic finite time, and remains at 0 once it reaches it.

Proof : The SDE verifies the condition that both the drift function and the diffusion function are Liptschitz with at most linear growth ; therefore the SDE has a unique strong solution. Let us denote by τ the potential hitting time at 0.
The proof that $\tau < 0$ almost surely involves the knowledge that a CIR process is a time-changed squared Bessel process [see e.g. Revuz and Yor (1999)].
Once the solution hits 0, it remains at 0 thereafter, as a consequence of the uniqueness of the solution from that date on. □

It is also useful to recall the link between the continuous time CIR process and the discrete time autoregressive gamma process [ARG, see e.g. Gouriéroux and Jasiak (2006)], both of which are affine processes. Let us first give the definition of an ARG process.

Definition B.1. A random variable F follows a noncentered gamma distribution $\tilde{\gamma}(\delta, \beta, c)$ if and only if there exists a Poisson variable with parameter β, $Z \sim \mathcal{P}(\beta)$ such that :

$$F \sim c\gamma(\delta + Z),$$

where γ is the standard gamma distribution.

Definition B.2. A process $(F_t, t = 1, 2, ...)$ is an autoregressive gamma process (of order 1, with constant coefficients δ, β and c) if the conditional distribution of F_t given F_{t-1} is $\tilde{\gamma}(\delta, \beta F_{t-1}, c)$.

Lemma B.4. The CIR process defined by (III-16) is such that the discrete time process $(F_t, t = 1, 2...T)$ is an autoregressive gamma (ARG) process with coefficients $\delta = 0$, $c = \sigma^2 \frac{1-e^{-m}}{2m}$, $\beta = e^{-m}/c$. The ARG process is positive before the hitting time τ of the CIR process, and remains null afterwards.

Proof : See Gouriéroux and Jasiak (2006). □

Since $\delta = 0$, and Z is a Poisson variable, there is a non zero probability that this ARG process hits zero at each date t. But this probability is negligible when the value of the process is large, or when σ is small.

Appendix B.6 Simulating the unobserved paths

The methodology used in this section is similar to that by Duffie et al. (2009). For simplicity, let us denote the unobserved frailty process by $F = (F_1, F_2, ..., F_T)$ where T is the number of values of the dynamic factor process F.

B.6.1 The Gibbs sampler

In order to generate samples of the path $(F_1, ..., F_T)$ conditional both on the value of parameter θ and all the observations Y_2, we can define a Markov chain $M = (M_k) = \Big((F_{1,k}, F_{2,k}, ..., F_{T,k})\Big)$ with values on the T-dimensional domain $(\mathbb{R}^+)^T$. If this multivariate chain is stationary with stationary distribution $F \mid \theta, Y_2$, then for large k, M_k will correspond to a drawing from this distribution. Such a chain can be constructed by the multi-step Gibbs sampler. The following theorem explains its principle :

Theorem B.1 (Hammersley and Clifford (1968))**.** Let $(X_1, X_2, ..., X_p)$ be a distribution with joint density function $f(x_1, x_2, ..., x_p)$ then for all $(\xi_1, \xi_2, ..., \xi_p) \in supp(f)$, we have :

$$f(x_1, ..., x_p) = \prod_{i=1}^{p} \frac{f_{(-j)(x_j | x_1, ...x_{j-1}, \xi_{j+1}, ..., \xi_p)}}{f_{(-j)(\xi_j | x_1, ...x_{j-1}, \xi_{j+1}, ..., \xi_p)}},$$

where $f_{(-j)}(\cdot \mid x_1, ...x_{j-1}, x_{j+1}, ..., x_p)$ is the conditional distribution function of X_j given all other X_i for $i \neq j$. These conditional distributions are called full conditional and the theorem states that they fully determine the joint distribution.

Now let us explain how to define the multivariate Markov chain (M_k) :

i) Initialize the value $M_1 = (F_{1,1}, F_{2,1}, ..., F_{T,1})$. For instance we set $F_{t,1} = \exp(-m(t-1))$ for all $t = 1, ..., T$, which corresponds to a deterministic factor as in the model with deterministic factor.

ii) Given the $k-$th value of the chain $M_k = (F_{1,k}, F_{2,k}, ..., F_{T,k})$, draw recursively the values

$F_{1,k+1}, F_{2,k+1}, ..., F_{T,k+1}$ in the following conditional univariate distributions :

$$
\begin{aligned}
& F_{1,k+1} \mid F_{2,k}, ..., F_{T,k}, Y_2, \theta \\
& F_{2,k+1} \mid F_{1,k+1}, F_{3,k} ..., F_{T,k}, Y_2, \theta \\
& F_{3,k+1} \mid F_{1,k+1}, F_{2,k+1}, F_{4,k}, ..., F_{T,k}, Y_2, \theta \\
& \dots \\
& F_{T,k+1} \mid F_{1,k+1}, F_{2,k+1}, ..., F_{T-1,k+1}, Y_2, \theta
\end{aligned}
\tag{B-8}
$$

In other words, the chain is updated component by component, by drawing at each iteration in a univariate distribution of the $F_{t,k+1}$ conditional on the parameter θ, the current values of other components of F, as well as the observation Y_2. This approach above cannot be used directly since the conditional distributions do not have forms appropriate for such a drawing [34]. Indeed, only the p.d.f. is easily calculable, up to a multiple constant (see below). But samples from these distributions can be approximated by means of the Metropolis-Hasting algorithm. This is explained in the next subsection.

iii) Store the new value of the chain $M_{k+1} = (F_{1,k+1}, F_{2,k+1}, ..., F_{T,k+1})$ and return to step *ii*).

To generate each of the T distributions given by (B-8), we employ a Metropolis-Hasting algorithm. Thus to generate the first K values of the Markov chain (M_k), we need to use KT times the Metropolis-Hasting algorithm.

B.6.2 The Metropolis-Hasting algorithm

Now let us explain the Metropolis-Hasting algorithm we used in the previous step *ii*). For each t, we should draw from the distribution

$$F_{t,k+1} \mid F_{1,k+1}, ..., F_{t-1,k+1}, F_{t+1,k}, ..., F_{T,k}, Y_2, \theta,$$

or $F_t \mid F_{(-t)}, Y_2, \theta$ for simplicity, where $F_{(-t)}$ denotes the vector $(F_1, F_2, ..., F_{t-1}, F_{t+1}, ..., F_T)$. Let us first explain how to calculate the p.d.f. of this conditional distribution.

34. More precisely, the corresponding cumulative distribution function, which should be used when simulating from a given distribution, cannot be calculated.

Using the same proof as in Duffie et al. (2009), especially the Markov property of F, we have :

$$p(F_t \mid F_{(-t)}, Y_2, \theta) \propto \mathcal{L}(\theta \mid Y_2, F) p(F_t \mid F_{t-1}, \theta) p(F_t \mid F_{t+1}, \theta). \tag{B-9}$$

The right hand side is the product of two terms. The first is $\mathcal{L}(\theta \mid Y_2, F)$, which is the likelihood of the lifetime data with given values F of the frailty process, that is,

$$\mathcal{L}(\theta \mid Y_2, F) = \exp \sum_{t_0} \Big\{ \sum_{i \in \eta_{t_0}^u} \log f_2(y_{2,i}, t_0, F) + \sum_{i \in \eta_{t_0}^c} \log S_2(y_{2,i}, t_0, F) \Big\},$$

where the expressions of $f_2(y_{2,i}, t_0, F)$ and $S_2(y_{2,i}, t_0, F)$ are the integrand in the right hand side of equations (B-4) and (B-5), respectively. This can be calculated for given values of θ and F. The second term is $p(F_t \mid F_{t-1}, \theta) p(F_t \mid F_{t+1}, \theta)$, which involves only the one-step transition density of the process (F_t) (given θ). Since it is an autoregressive gamma process, this transition density can be calculated in an exact way. Therefore the second term is equally easy to calculate. Thus the density function given by (B-9) can be evaluated at each point up to a multiple constant. Instead of drawing directly from this distribution, we can define an auxiliary univariate Markov chain denoted by $(F_{t,k}^{(n)}, n = 1, 2, ...)$, or $F_t^{(n)}$ for simplicity. This chain is also stationary and its stationary distribution is given by (B-9). Thus we can approximate $F_{t,k+1}$ by $F_t^{(n)}$ for a large value of n. The transition rule of this Markov chain $F_t^{(n)}$ is described as follows :

1. Initialize the chain by setting $F_t^{(1)} = 1$.
2. For $n = 2, 3, ...$, draw a candidate from a proposal distribution, for instance, we can choose the log-normal distribution [35] :

$$f \sim F_t^{(n-1)} \mathcal{N}(0, \sigma),$$

   where the standard deviation of the proposal density is chosen arbitrarily, say, $\sigma_p = 0.01$.

3. Compute

$$\alpha = \frac{p(F_t = f \mid F_{(-t)}, Y_2, \theta)}{p(F_t = F_t^{(n-1)} \mid F_{(-t)}, Y_2, \theta)}, \tag{B-10}$$

   where both the numerator and the denominator can be calculated by equation (B-9).

35. This choice is mainly motivated by simplicity reasons. Indeed it allows for a symmetric conditional density since $p(f|F_t^{(n-1)}) = p(F_t^{(n-1)}|f)$, so that there is no need to compute the ratio $\frac{p(f|F_t^{(n-1)})}{p(F_t^{(n-1)}|f)}$. Besides, we should use a positive distribution, (since the factor F is nonnegative), which is the case for the log-normal distribution.

4. Draw a uniform variable $u \sim U([0,1])$ and set the $n-$th value F_t^n by the following rule : [36]

$$F_t^{(n)} = \begin{cases} f, & \text{if } u < \alpha \\ F_t^{(n-1)}, & \text{otherwise} \end{cases}$$

To ensure the convergence of this univariate Markov chain to its stationary distribution (B-9), we take, say, the 300 th value of the chain as a sample from this distribution, which is used in step *ii*) of the Gibbs sampling algorithm.

Appendix B.7 Identification proof of Proposition III.3

Remind that with a deterministic exponential factor, the age-cohort and age-calendar time models are equivalent and that the survivor function for cohort t_0 is given by :

$$\begin{aligned} S_2(y_2,t_0) = \int_0^{y_2} [a_1(x) + \tilde{b}_1(x)F_{t_0}] \exp\Big(- \int_0^x [a_1(s) + \tilde{b}_1(s)F_{t_0}]ds \\ - \int_0^x [a_2(s) + \tilde{b}_2(s)F_{t_0}]ds - \int_x^{y_{2,i}} [a_3(s) + \tilde{b}_3(s)F_{t_0}]ds\Big)dx \\ + \exp\Big(- \int_0^{y_2} [a_1(x) + \tilde{b}_1(x)F_{t_0}]dx - \int_0^{y_2} [a_2(x) + \tilde{b}_2(x)F_{t_0}]dx\Big). \end{aligned} \tag{B-11}$$

B.7.1 Identification of m.

When $y_2 \to 0$, we have, for $t_1 \neq t_0 \neq t_2 \neq t_1$,

$$\lim_{y_2 \to 0} \frac{\lambda(y_2,t_2) - \lambda(y_2,t_0)}{\lambda(y_2,t_1) - \lambda(y_2,t_0)} = \frac{e^{-mt_2} - e^{-mt_0}}{e^{-mt_1} - e^{-mt_0}}.$$

Since the LHS in the equation above is observable, m is point identified. Note that the identification assertion remains valid even for a general functional parameter (F_t) without the exponential specification, under the limiting longevity assumption $\lim_{t\to\infty} F_t = 0$. Indeed, under this assumption the ratio $\frac{F_{t_2}-1}{F_{t_1}-1}$ is identified, where we remind that $F_{t_0} = 1$. If there is another path (F'_t) such that

$$\frac{F_{t_2} - 1}{F_{t_1} - 1} = \frac{F'_{t_2} - 1}{F'_{t_1} - 1},$$

36. The equation B4 in Duffie et al. (2009)[Appendix C] is not correct since their α does not depend on the factor $p(F_t \mid F_{t-1}, \theta)p(F_t \mid F_{t+1}, \theta)$.

then $\frac{F_{t_2}-1}{F'_{t_2}-1} = \frac{F_{t_1}-1}{F'_{t_1}-1}$ is equal to a constant that does not depend on t_2, t_1. Let t_2 go to infinity, by using the limiting condition $\lim_{t\to\infty} F_t = 0$, we deduce that this constant equals 1. Thus the path of the process (F_t) is nonparametrically identified. Moreover, the following identification of functional parameters remains valid for a general form of (F_t) and the age-cohort specification, but not the age-calendar time model, except with the exponential specification. See also the discussion in Section 4.2.2.

B.7.2 Identification of functional parameters $a_1, a_2, a_3, b_1, b_2, b_3$.

Under the assumption that all functions[37] $a_1, b_1, a_2, b_2, a_3, b_3$ are continuous, the conditional survivor function $S(y_2|t_0) = S(y_2|F)$ is an analytic function of F for a given y. Therefore it is equivalent to know this function or to know all its derivatives for any pre-specified t_0. These derivatives are simpler to deal with, especially if $t_0 = \infty$; equivalently we look at the derivative at $F = 0$. The case $t_0 < \infty$ is similar[38]. Thus we obtain, at order 0,

$$\int_0^{y_2} a_1(x) e^{-A_1(x)-A_2(x)-A_3(y_2)+A_3(x)} dx + e^{-A_1(y_2)-A_2(y_2)} = S(y_2, F = 0), \tag{B-12}$$

and at each order $n \geq 1$,

$$\begin{aligned}\int_0^{y_2} e^{-A_1(x)-A_2(x)-A_3(y_2)+A_3(x)} \Bigg(& \frac{a_1(x)(-1)^n}{n!} \big[B_1(x) + B_2(x) + B_3(y_2) - B_3(x)\big]^n \\ & + \frac{b_1(x)(-1)^{n-1}}{(n-1)!} \big[B_1(x) + B_2(x) + B_3(y_2) - B_3(x)\big]^{n-1} \Bigg) dx \\ & + e^{-A_1(y_2)-A_2(y_2)} \frac{(-1)^n}{n!} \Big[B_1(y_2) + B_2(y_2)\Big]^n = (-1)^n \frac{\partial S}{\partial F}(y_2, F = 0),\end{aligned} \tag{B-13}$$

for all $y_2 \in [0, T]$, where the capital letters denote the cumulative integrals of the corresponding lower case functions.

Except in some special cases, one expects that (B-12) and (B-13) give a non degenerated infinite system of functional equations that $a_1, b_1, a_2, b_2, a_3, b_3$ should satisfy. This raises hopes that the solution to such a system is generically unique. Let us first look at Case 1 in Proposition III.3.

37. Strictly speaking, the functional parameters are $a_1, a_2, a_3, \tilde{b_1}, \tilde{b_2}, \tilde{b_3}$. For ease of exposure, we omit the tilde symbol on b_1, b_2, b_3 for the rest of this section.

38. If $t_0 < \infty$, we should look at the sequence of derivatives of the function $S_2(y_2|F_{t_0})$ for any given y_2 at the point $F_{t_0} \neq 0$. Their expressions are more complicated than at point $F_{t_0} = 0$.

Case 1 (global identification). If $b_1 + b_2 = b_3$, the $n-$th equation becomes :

$$B_3(y_2)^{n-1} \int_0^{y_2} e^{-A_1(x)-A_2(x)-A_3(y_2)+A_3(x)} \left(\frac{a_1(x)(-1)^n}{n!} B_3(y_2) - \frac{b_1(x)(-1)^{n-1}}{(n-1)!} \right) dx$$
$$+ e^{-A_1(y_2)-A_2(y_2)} \frac{(-1)^n}{n!} B_3(y_2)^n = (-1)^n \frac{\partial S}{\partial F}(y_2, F = 0).$$

For $y > 0$, $B_3(y_2) > 0$, and large n, the LHS of the equation above is equivalent to :

$$B_3(y_2)^{n-1} \frac{(-1)^n}{n!} \int_0^{y_2} e^{-A_1(x)-A_2(x)-A_3(y_2)+A_3(x)} b_1(x) dx.$$

Therefore $B_3(y_2)$ is globally identified [39], as well as the constant (in n) $\int_0^{y_2} e^{-A_1(x)-A_2(x)-A_3(y_2)+A_3(x)} b_1(x) dx$. Then by suppressing this dominating term, the LHS of the previous $n-$th equation reduces to the LHS in (B-12). Thus the infinite system reduces to only three independent equations and the model is not identified.

Case 2 (global identification). There exists constants $c, c' > 0$ such that $b_1 + b_2 - b_3 \geq c$ and $|b_2 - b_3| > c'$. For expository purpose let us introduce the following functions :

$$C(y) = e^{-A_1(y)-A_2(y)},$$
$$D(y) = B_1(y) + B_2(y),$$
$$f_n(y) = \int_0^y e^{-A_1-A_2-A_3(y)+A_3} b_1 \Big[B_1 + B_2 + B_3(y) - B_3 \Big]^n dx,$$
$$g_n(y) = \int_0^y e^{-A_1-A_2-A_3(y)+A_3} a_1 \Big[B_1 + B_2 + B_3(y) - B_3 \Big]^n dx.$$

39. By global identification, we refer to the standard definition of identification, that is, a function is identified if at any point y_2, the value of this function is uniquely determined. This notion has to be distinguished from the concept of local (nonparametric) identification, as in Chen et al. (2014), detailed later on in the proof.

Since $B_1(x)+B_2(x)+B_3(y)-B_3(x)$ is positive, increasing in x, and the term $e^{-A_1-A_2-A_3(y)+A_3}b_1$ is positive and bounded, we can prove that [40] :

$$(n+1)f_n(y) \sim \frac{e^{-A_1(y)-A_2(y)}b_1(y)}{b_1(y)+b_2(y)-b_3(y)}\Big[B_1(y)+B_2(y)\Big]^{n+1} \tag{B-14}$$

when n goes to infinity and

$$(n+1)g_n(y) \sim \frac{e^{-A_1(y)-A_2(y)}a_1(y)}{b_1(y)+b_2(y)-b_3(y)}\Big[B_1(y)+B_2(y)\Big]^{n+1}.$$

Then we can study the behavior of the LHS of (B-13). We have :

$$\begin{aligned}(-1)^n\frac{\partial S}{\partial F}(y_2,F=0) &= \frac{(-1)^n}{n!}g_n(y)+\frac{(-1)^{n-1}}{(n-1)!}f_{n-1}(y)+e^{-A_1(y_2)-A_2(y_2)}\frac{(-1)^n}{n!}\Big[B_1(y_2)+B_2(y_2)\Big]^n \\ &\sim e^{-A_1(y_2)-A_2(y_2)}\frac{(-1)^n}{n!}\Big[B_1(y_2)+B_2(y_2)\Big]^n\Big(1-\frac{b_1(y_2)}{b_1(y_2)+b_2(y_2)-b_3(y_2)}\Big)\end{aligned}$$

provided that b_2-b_3 is never null. Then $B_1(y)+B_2(y)$ is globally identified, as well as the function $\frac{e^{-A_1(y)-A_2(y)}\left(b_2(y)-b_3(y)\right)}{b_1(y)+b_2(y)-b_3(y)}$.

Case 3 (global identification). Similarly, if there exists $d>0$ such that $b_1+b_2-b_3\leq -d$, then $(n+1)f_n(y) \sim \frac{e^{-A_3(y)}b_1(0)}{b_1(0)+b_2(0)-b_3(0)}B_3(y)^n$. $B_3(y)$ is globally identified, as well as $e^{-A_3(y)}$, up to an additive constant. Since $A_1(0)=0$, the constant is uniquely determined. Therefore A_3 is identified as well.

Cases 2,3 (local identification). Let us finally prove that the other functions are locally identified. We do this by following Chen et al. (2014), who give the definition of local identification on a functional space. Roughly speaking, a function h is locally identified at h_0, if h_0 is the unique solution to a certain system of equations when the unknown function is restricted to be in a

40. Intuitively, when n becomes large, the contribution of the integrand at a point x that is away from y_2 is negligible since $\Big[B_1(x)+B_2(x)+B_3(y)-B_3(x)\Big]^n$ is much smaller than $\Big[B_1(y)+B_2(y)\Big]^n$. Thus the asymptotic behavior of this integral depends only on the behavior of the integrand in a neighbourhood of point y. To get another informal explanation of this result, we can use the integration by parts :

$$\begin{aligned}(n+1)f_n(y) = &\frac{e^{-A_1(y)-A_2(y)}b_1(y)}{b_1(y)+b_2(y)-b_3(y)}\Big[B_1(y)+B_2(y)\Big]^{n+1}-\frac{e^{-A_3(y)}b_1(0)}{b_1(0)+b_2(0)-b_3(0)}B_3(y)^{n+1} \\ &-\int_0^y\frac{\partial}{\partial x}\Big(\frac{e^{-A_1-A_2-A_3(y)+A_3}b_1}{b_1+b_2-b_3}\Big)\Big[B_1+B_2+B_3(y)-B_3\Big]^{n+1}dx\end{aligned}$$

Since $\frac{B_1(y)+B_2(y)}{B_3(y)}>1$, the second term is negligible with respect to the first one; if the partial derivative in the third term exists and is bounded, then the third term is $O(f_n(y))$ when n goes to infinity. By rearranging this equation, we get the desired asymptotic equivalent. The formal proof of this result uses solely real analysis techniques and does not requires the existence of the partial derivative which is needed the integration by parts.

certain neighbourhood of h_0 [see Definition 1, Chen et al. (2014)]. In our case the neighborhood has to be defined on an appropriate functional space and we have to find a functional operator, whose Gâteaux derivative is non degenerated. This is the infinite dimensional analogue of the standard full rank condition for local identification of parametric models. As explained in Chen et al. (2014), on the contrary to the finite dimensional case where the rank condition is also sufficient, in an infinite dimensional space, this condition alone implies only a rather weak notion of local identification [see Theorem 2, Chen et al. (2014)].

For expository purpose, let us focus on Case 3. For Case 2, the calculations are slightly more complicated, but the principle stays the same.

Let us denote by $\mathcal{B} = \mathcal{C}([0,T])$ the space of all continuous functions on the age domain $[0,T]$, where T is a fixed constant, that is, we assume that the observations are only available up to a maximum age, say, $T = 110$. We have deliberately chosen a fixed upper bound [41] so that the functional space $\mathcal{B}$, topologized by the uniform norm $||f|| = \max_{t\in[0,T]} |f(t)|$, is a Banach space. This fixed upper bound is not restrictive since, if we can prove local identification for any given T, then we will have local identification on the whole age domain $[0,\infty[$. Also remind that on the space $\mathcal{B}$, all functions are bounded, and all positive functions are lower bounded by a positive constant. Under this framework, we have the following Lemma, which is a direct consequence of Theorem 2 in Chen et al. (2014) :

Lemma B.5. The functions (a_1, b_1, a_2, b_2) are locally identified in the sense of Theorem 2 in Chen et al. (2014) if the following four conditions are satisfied :

i) For each $n \geq 1$, the LHS of (B-13) is a continuous operator from $\mathcal{A} := \mathcal{B}^4$ to space $\mathcal{B}$, with the corresponding uniform topology for each space. These operators are denoted $m_n : \mathcal{A} \mapsto \mathcal{B}$.

ii) For each order $n \geq 1$, the operator m_n is Fréchet differentiable [see e.g. Chen et al. (2014) Equation 2.1]. For each element $\alpha \in \mathcal{A}$ we denote by $h \mapsto m_n'(h)$ the Fréchet derivative at point α, where h is the generic element of the space $\mathcal{A}$, $n \geq 0$. This derivative depends on the point $\alpha \in \mathcal{A}$, but we will omit the index α.

iii) The intersection of the null spaces $\cap_{n=1}^{\infty}$Ker m_n' is reduced to $\{\mathbf{0}\}$.

These conditions are quite intuitive. Condition *i)* is a regularity condition at both infinity (since $y_2 \leq T < \infty$) and zero (since the integrands are all bounded at zero). It excludes in

41. The assumption of a fixed upper bound for the observable attained age is compatible with the previous assumption $t_0 = \infty$, on the observed cohort.

particular mixed proportional hazard (MPH) models with heavy-tailed unobserved heterogeneity distribution and an intensity function that is equal to infinity at time zero [see e.g. Ridder (1990)]. Condition *ii*), that is the differentiability of these operators, is clearly satisfied, since each operator is a compounding of elementary (Gâteaux−) differentiable operators. Condition *iii*) is Assumption 1 in Chen et al. (2014), and is the infinite dimensional analogue of the full rank condition.

Let us give the proof of the lemma. For given (a_3, b_3) as well as path of (F_t), the survivor function $S(y|F)$ is a bivariate continuous function in arguments (y, F), that is $S(y|F) \in \mathcal{C}([0,T] \times [0,1])$ which is a Banach space. Denote by M the operator from $\mathcal{A}$ to $\mathcal{C}([0,T] \times [0,1])$, which maps the point (a_1, b_1, a_2, b_2) to the corresponding survivor function $S(y|F)$. Then by Chen et al. (2014), it suffices to prove that M', the Gâteaux derivative of M is nonsingular [42]. Because $S(y|F)$ (as well as its Gâteaux derivative) is analytical in F, $M'(y,F) = 0$ is equivalent to the derivatives of any order with respect to F being null functions. These derivatives are exactly [43] the sequence m'_n.

Let us finally check that Condition *iii*) is satisfied in our framework. The expression of $m'_n(h)$ at point (a_1, b_1, a_2, b_2), for any $h = (da_1, db_1, da_2, db_2) \in \mathcal{A}$, is the following :

$$
\begin{aligned}
&(-1)^n (n-1)! m'_n(h)(y) \\
&= \int_0^y e^{-A_1 - A_2 - A_3(y) + A_3} \frac{a_1}{n} \Big[B_1 + B_2 + B_3(y) - B_3 \Big]^n \Big[- dA_1 - dA_2 \Big] dx \\
&\quad + \int_0^y e^{-A_1 - A_2 - A_3(y) + A_3} \frac{da_1}{n} \Big[B_1 + B_2 + B_3(y) - B_3 \Big]^n dx \\
&\quad + \int_0^y e^{-A_1 - A_2 - A_3(y) + A_3} a_1 \Big[B_1 + B_2 + B_3(y) - B_3 \Big]^{n-1} \Big[dB_1 + dB_2 \Big] dx \\
&\quad - \int_0^y e^{-A_1 - A_2 - A_3(y) + A_3} b_1 \Big[B_1 + B_2 + B_3(y) - B_3 \Big]^{n-1} \Big[- dA_1 - dA_2 \Big] dx \\
&\quad - \int_0^y e^{-A_1 - A_2 - A_3(y) + A_3} db_1 \Big[B_1 + B_2 + B_3(y) - B_3 \Big]^{n-1} dx \\
&\quad - \int_0^y e^{-A_1 - A_2 - A_3(y) + A_3} (n-1) b_1 \Big[B_1 + B_2 + B_3(y) - B_3 \Big]^{n-2} \Big[dB_1 + dB_2 \Big] dx \\
&\quad + e^{-A_1(y) - A_2(y)} \frac{\Big[B_1(y) + B_2(y) \Big]^{n-1}}{n} \Bigg(- \Big[dA_1(y) + dA_2(y) \Big] \Big[B_1(y) + B_2(y) \Big] + n \Big[dB_1(y) + dB_2(y) \Big] \Bigg),
\end{aligned}
\tag{B-15}
$$

42. $M'(\big(da_1, db_1, da_2, db_2\big)\big)$ is a bivariate function in arguments y and F.

43. We have used the fact that it is equivalent to first take derivative with respect to F, then the Gâteaux derivative with respect to (a_1, b_1, a_2, b_2) or conversely.

where $dA_1, dA_2, ..$ are cumulative integral of the corresponding lower case functions. Let us explain this formula : lines 1-3 (resp. lines 4-6 and line 7) are the Gâteaux derivatives of the first (resp. second and third) term of the LHS of (B-13).

Assume now that $m_n'(h) = 0$ for a certain function $h = (da_1, da_2, db_1, db_2)$ and for all $n \geq 0$. Similarly as (B-14), when n goes to infinity, $m_n'(h)$ is equivalent to :

$$-\frac{C(y)}{b_1(0) + b_2(0) - b_3(0)} \Big(b_1(y)dB_1(y) - (b_2(y) - b_3(y))dB_2(y) \Big) B_3^{n-1}(y)$$

provided that this term is non null. Thus we should have :

$$b_1(y)dB_1(y) - (b_2(y) - b_3(y))dB_2(y) = 0 \tag{B-16}$$

for all y. Then similarly, $m_n'(h)$ is equivalent to :

$$-\frac{C(y)}{(n-1)\Big[b_1(0) + b_2(0) - b_3(0)\Big]} \Big(a_1(y)dB_1(y) + a_2(y)dB_2(y) + b_1(y)dA_1(y) - (b_2(y) - b_3(y))dA_2(y) \Big) B_3^n(y),$$

provided that this term is non null. Therefore :

$$a_1(y)dB_1(y) + a_1(y)dB_2(y) + b_1(y)dA_1(y) - (b_2(y) - b_3(y))dA_2(y) = 0. \tag{B-17}$$

Similarly, we have

$$a_1(y)dA_1(y) - a_1(y)dA_2(y) = 0, \tag{B-18}$$

and finally

$$-\frac{C(y)D^n(y)}{n} dA_2(y) + C(y)D^{n-1}(y)dA_1(y) = 0, \quad \forall n. \tag{B-19}$$

Combining (C-7) to (B-19) we can get $dA_1 = dA_2 = 0$, and then we have $dB_1 = dB_2 = 0$, except when :

$$\frac{b_3 - b_2}{b_1} = \frac{a_1}{a_1} = 1,$$

which is not allowed since Case 3 assumes $b_1 + b_2 - b_3 < 0$. Thus Condition *iii*) above is satisfied.

Chapitre IV

Large Duration Asymptotics in Bivariate Survival Models with Unobserved Heterogeneity

Abstract

A major risk for pension funds is due to the pensioners living a very long time, called advanced age survivors, and this risk increases following the general evolution of human lifetimes, that is the longevity phenomenon. This paper focuses on such joint advanced age survivors in the framework of bivariate survival models with bivariate unobserved heterogeneity. We first give minimal conditions to ensure that the bivariate heterogeneity still exists among advanced age survivors. Then, under these conditions, we derive the necessary form of the joint duration distribution among advanced age survivors. This large duration asymptotics depends on two functional parameters, which characterize the survivor probability, and the joint dependence between the two survival variables given survival, respectively.

These large duration asymptotics of the survival variables are closely related with the behavior near zero of the heterogeneity distribution. More precisely, under the same conditions, the heterogeneity distribution among survivors converges to a limit semi-parametric distribution. This generalizes the univariate result derived in Abbring and van den Berg (2007).

Keywords : Dependent Competing Risks, Unobserved Heterogeneity, Regular Variation, Non-parametric Identification, Human Longevity.

IV.1 Introduction

A major risk for pension funds is due to the pensioners living a very long time, called advanced age survivors, and this risk increases following the general evolution of human lifetimes, that is the longevity phenomenon. This paper focuses on such advanced age survivors in the framework of bivariate survival models with bivariate unobserved heterogeneity. We first give minimal conditions to ensure that the bivariate heterogeneity still exists among advanced age survivors. Then, under these conditions, we derive the necessary form of the duration distribution among the advanced age survivors. This large duration asymptotics depends on a functional parameter, which characterize the joint dependence between the two survival variables among advanced age survivors, respectively.

The paper is structured as follows. I review in Section 2 the large duration asymptotic results for univariate survival models with unobserved heterogeneity derived in Abbring and van den Berg (2007). Loosely speaking, after an appropriate time change, the heterogeneity distribution among advanced age survivors is asymptotically gamma and the duration distribution asymptotically Pareto, under a condition of regular variation at zero of the initial heterogeneity distribution. Section 3 extends this analysis to bivariate survival variables with bivariate heterogeneity. I derive conditions for the heterogeneity distribution among the survivors to converge, and study properties of the limit distribution. I also provide an alternative interpretation of the convergence result in terms of the asymptotic behavior of the survival variables. Section 4 discusses the identification of asymptotic parameters from large duration samples. Section 5 concludes. Proofs and technical lemmas are gathered in Appendices.

IV.2 Advanced age survivors in univariate models

Let us denote by T the survival variable and assume that its conditional intensity at age (duration) t, conditional on the observed individual characteristics z and the unobserved individual heterogeneity U, respectively, is proportional to U :

$$\theta(t|z,U) := \lim_{\mathrm{d}t\to 0} \frac{1}{\mathrm{d}t}\mathbb{P}[T < t + \mathrm{d}t | T > t, z, U] = \lambda(t,z)U.$$

The function λ is the baseline intensity function and its cumulative integral with respect to t is denoted Λ. This specification nests the mixed proportional hazard (MPH) model [see e.g. Elbers

and Ridder (1982)], where $\lambda(t,z) = \lambda(t)\phi(z)$, as well as the generalized accelerated failure time model $\lambda(t,z) = \lambda(t\phi(z))$ [see e.g. Ridder (1990)]. The observable characteristics z are temporarily omitted for expository purpose.

Under this specification, the survivor function is $\mathbb{P}[T > t] = \mathbb{E}[e^{-\Lambda(t)U}] = \int e^{-\Lambda(t)u}\mathrm{d}F(u)$, where F is the cdf of U, and the hazard function is :

$$h(t) := -\frac{\mathrm{d}}{\mathrm{d}t}\log\mathbb{P}[T > t] = \frac{\int \lambda(t)ue^{-\Lambda(t)u}\mathrm{d}F(u)}{\int e^{-\Lambda(t)u}\mathrm{d}F(u)} = \lambda(t)\mathbb{E}[U|T > t], \tag{IV-1}$$

where $\frac{e^{-\Lambda(t)u}}{\int e^{-\Lambda(t)v}\mathrm{d}F(v)}\mathrm{d}F(u)$ is the conditional distribution of U given $T > t$, namely the heterogeneity distribution among the survivors at time t.

Throughout this paper, we consider only non defective individuals and assume :

Assumption IV.1. The cumulative intensity $\lim_{t\to\infty}\Lambda(t) = \infty$, and U has no point mass at zero.

From a large duration point of view, the quantities of interest are the conditional distribution of T given $T > t$, as well as the survival probability $\mathbb{P}[T > t]$, for large t.

IV.2.1 Conditional distribution of T given $T > t$

By Bayes' formula we have :

$$\mathbb{P}[T > t+\tau|T > t] = \frac{\int e^{-\Lambda(t+\tau)u}\mathrm{d}F(u)}{\int e^{-\Lambda(t)v}\mathrm{d}F(v)} = \int e^{-[\Lambda(t+\tau)-\Lambda(t)]u}\frac{e^{-\Lambda(t)u}}{\int e^{-\Lambda(t)v}\mathrm{d}F(v)}\mathrm{d}F(u),$$

where $\frac{e^{-\Lambda(t)u}}{\int e^{-\Lambda(t)v}\mathrm{d}F(v)}\mathrm{d}F(u)$ is the heterogeneity distribution among survivors at time t. Since the term $e^{-\Lambda(t)u}$ suggests the scale change in heterogeneity $U_t^* = \Lambda(t)U$, let us denote by F_t the distribution of $U_t^* = \Lambda(t)U$ given $T > t$. This leads to another time-change for the duration variable T :

$$X_t = \frac{\Lambda(T)}{\Lambda(t)} - 1, \quad \text{say}, \tag{IV-2}$$

when $T > t$. This is both a change of time origin and an increasing non-linear change of time unit [1]. Under these variable changes, we get :

$$\mathbb{P}[T > t+\tau|T > t] = \mathbb{P}\Big[X_t > \frac{\Lambda(t+\tau)}{\Lambda(t)} - 1|T > t\Big] = \int e^{-\left[\frac{\Lambda(t+\tau)}{\Lambda(t)}-1\right]u^*}\mathrm{d}F_t(u^*). \tag{IV-3}$$

1. The time change Λ is common in the literature [see e.g. Horowitz (1999)] : without the conditioning on $T > t$, but conditioning on U, it is well known that $\Lambda(T)U$ follows a unit exponential distribution. In our case we integrate U out and condition with respect to $T > t$.

Thus, if F_t converges, then the distribution of X_t given $T > t$ converges as well. The following theorem provides the condition of this convergence in terms of the behavior at zero of frailty U.

Theorem IV.1 (Abbring and van den Berg (2007))**.** Under Assumption 1, the distribution of $\Lambda(t)U$ given $T > t$ converges to a non degenerate distribution, when t goes to infinity, if and only if the cdf F is regularly varying at zero (RV_0) with an index $\alpha \geq 0$, namely for all $a \in [0, 1]$,

$$\lim_{x \to 0} \mathbb{P}\Big[U < ax | U < x\Big] = \lim_{x \to 0} \frac{F(ax)}{F(x)} = a^{\alpha}.$$

In this case, the limit distribution is necessarily gamma [2] $\gamma(\alpha, 1)$.

Properties of regularly varying functions are gathered in Appendix C.1. In particular, it is recalled that the limit is necessarily of the form $a^{\alpha} = \lim_{t \to \infty} \frac{F(ax)}{F(x)}$. Then $F(x)$ can be alternatively written as $F(x) = x^{\alpha} L(x)$, where L is slowly varying at zero, that is, $\lim_{x \to 0} \frac{L(ax)}{L(x)} = 1$ for all $a > 0$. Roughly speaking, such a function varies very slowly (see Lemma C.2) for small x and under certain circumstances can be approximately regarded as a constant. Thus regularly varying functions have a "quasi" power decreasing rate near zero.

From Theorem 1 and Equation (IV-3), we deduce the following property :

Property IV.1. Under Assumption 1 and the regular variation assumption of F at zero, we have, for each $x \geq 0$,

$$\mathbb{P}[X_t > x | T > t] \to \int e^{-xu^*} \mathrm{d}F_{\alpha,1}(u^*) = \frac{1}{(1+x)^{\alpha}},$$

where $F_{\alpha,1}$ is the cdf of a gamma distribution $\gamma(\alpha, 1)$. Thus the distribution of X_t given $T > t$ converges to a Pareto distribution.

IV.2.2 Marginal tail of T

Let us now consider the tail properties of the marginal distribution of T.

Property IV.2. Under the regular variation assumption of F at zero, we have :

1. $\mathbb{P}[T > t] = \frac{L^*(\Lambda(t))}{\Lambda^{\alpha}(t)}$, where L^* is slowly varying at infinity : $\lim_{t \to \infty} \frac{L^*(at)}{L^*(t)} = 1$ for all $a > 0$.
2. $h(t) \sim \alpha \frac{\lambda(t)}{\Lambda(t)} = \alpha \frac{\mathrm{d}}{\mathrm{d}t} \log \Lambda(t) = -\frac{\mathrm{d}}{\mathrm{d}t} \log \left[\frac{1}{\Lambda^{\alpha}(t)}\right]$, when t goes to infinity.

2. The distribution $\gamma(\alpha, 1)$ is defined by $\mathrm{d}F_{\alpha,1}(x) = \frac{1}{\Gamma(\alpha)} x^{\alpha} e^{-x} \mathrm{d}x$.

Proof : We have $\mathbb{P}[T > t] = \mathbb{E}[e^{-\Lambda(t)F}]$. Thus by Theorem C.1 (see Appendix C.1), we get Property 2.1). The second part of the property is a direct consequence of Theorem IV.1 and equation (IV-1). □

Thus the marginal tail of T is characterized by the scalar α, and the asymptotic behavior of Λ, via $\Lambda^\alpha(t)$ and its derivative.

IV.2.3 Illustration

Let us consider a simple case in which the conditional distribution of X_t given $T > t$ has a closed form for any t.

Example IV.1 (gamma frailty)**.** Assume that U initially follows a gamma distribution, $\gamma(\alpha, c)$, which is $RV_0(\alpha)$. Then we have :

$$\mathbb{P}[T > t + \tau | T > t] = \Big[1 + c\frac{\Lambda(t+\tau) - \Lambda(t)}{1 + c\Lambda(t)}\Big]^{-\alpha}.$$

Therefore, for any t, the distribution of $X_t = c\dfrac{\Lambda(T) - \Lambda(t)}{1 + c\Lambda(t)}$ given $T > t$ is a Pareto distribution with parameter α. Since $\lim_{t\to\infty} \Lambda(t) = \infty$, we get, for large t, $X_t \sim \frac{\Lambda(T)}{\Lambda(t)} - 1$, which is the result in Property 1. Moreover,

$$\mathbb{P}[T > t] = \Big[1 + c\Lambda(t)\Big]^{-\alpha} = \Lambda^{-\alpha}(t)\Big[c + \frac{1}{\Lambda(t)}\Big]^{-\alpha},$$

and the slowly varying function $L(t) = \Big[c + \frac{1}{\Lambda(t)}\Big]^{-\alpha}$ converges to a constant.

Let us illustrate, in Figure 1, the convergence of the heterogeneity distribution among survivors to the limiting distribution, when the initial distribution is regularly varying at zero, but non gamma. We set the baseline intensity as $\lambda(t) = 1$, and the cdf of U as $F(x) = 0.4\gamma_{1.5,2}(x) + 0.6\mathbb{1}_{x\geq 5}\gamma_{1.5,2}(x-5)$. In other words, U is a mixture of two gamma type variables $U_1 \sim \gamma(1.5, 2)$ and $U_2 \sim 5 + \gamma(1.5, 2)$, the latter one has a support which does not include 0. This distribution is regularly varying [3] with index $\alpha = 1.5$.

3. Indeed, the term corresponding to the second component has no impact on the regular variation behavior at zero, since for all $x < 5$, $0.6\mathbb{1}_{x\geq 5}\gamma_{1.5,2}(x-5) = 0$.

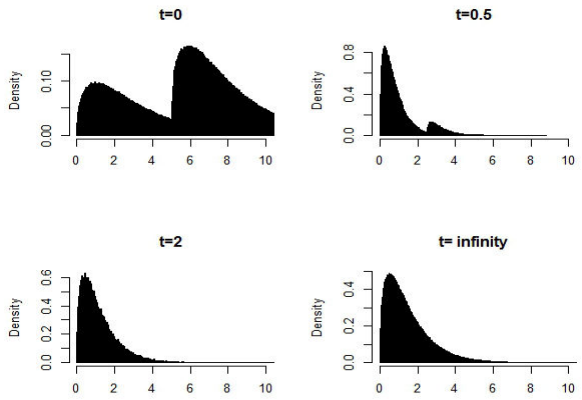


FIGURE IV-1: Evolution of the distribution of $\Lambda(t)U$ given $T > t$. The initial distribution has two modes ; as t increases, the second mode progressively wears off and the heterogeneity distribution among survivors converges to a gamma distribution, which has only one mode.

IV.3 Advanced age survivors in bivariate survival models

Let us now extend the analysis to bivariate survival models. For each individual, we denote by T_1 and T_2 the two event times. An individual satisfying $T_1 > t, T_2 > t$, when t is large, is called an advanced age survivor. Let us first provide potential applications with the corresponding definition of the advanced age survivors in each example.

i) **Competing risks model.** This is the standard example of an individual with two possible causes of death 1 and 2, say.

— T_1 is the potential time of death due to cause 1 ;

— T_2 is the potential time of death due to cause 2.

In this case, the survival variables T_1 and T_2 are latent : at most we can observe the time of failure $\min(T_1, T_2)$, as well as the cause of the failure, that is the indicator variable $\mathbb{1}_{T_1<T_2}$. The event $T_1 > t, T_2 > t$ means that the individual is still alive at time t.

ii) **Semi-competing risks model.** In this case, the individual can experience either a non-terminal event, followed by a terminal event (death), or he/she can also only experience the terminal event. For instance,

— T_1 is the potential time of entering into (permanent) long-term care (LTC) ;

— T_2 is the time of death.

In this case, $T_1 > t, T_2 > t$ means that the individual is still alive and autonomous at time t.

3) **Complete observations.** This situation arises when we consider the lifetimes of a couple, where

— T_1 the time of death of the husband ;

— T_2 the time of death of the wife.

In this example, an individual stands for a couple and the event $T_1 > t, T_2 > t$ means that both spouses are alive at time t.

By analogy with the univariate model in Section 2, we consider a bivariate model with proportional heterogeneity. More precisely we assume that the survival variables T_1 and T_2 are independent conditional on the unobservable individual heterogeneity (U, V), that their conditional intensities are :

$$\begin{aligned} \theta_1(t|U,V,T_1>t) &:= \lim_{\mathrm{d}t\to 0} \frac{1}{\mathrm{d}t}\mathbb{P}[T_1 < t+\mathrm{d}t | T_1 > t, T_2 > t, U, V] = \lambda_1(t)U, \\ \theta_2(t|U,V,T_2>t) &:= \lim_{\mathrm{d}t\to 0} \frac{1}{\mathrm{d}t}\mathbb{P}[T_2 < t+\mathrm{d}t | T_1 > t, T_2 > t, U, V] = \lambda_2(t)V, \end{aligned} \tag{IV-4}$$

where λ_1, λ_2 are baseline hazard functions. In this bivariate model, the heterogeneities U and V are not necessarily independent. They capture the spurious duration dependencies and, more importantly, the risk correlation between T_1 and T_2. Similarly, we can define the unconditional event-specific hazard functions by :

$$h_1(t) := \lim_{\mathrm{d}t\to 0} \frac{1}{\mathrm{d}t}\mathbb{P}[t < \min(T_1, T_2) < t+\mathrm{d}t, T_1 < T_2 \mid \min(T_1, T_2) > t], \text{ say.}$$

Below, we are interested in the joint distribution of (T_1, T_2) given $T_1 > t, T_2 > t$, as well as in the survival probability $\mathbb{P}[\min(T_1, T_2) > t]$, for large t.

IV.3.1 Asymptotically competing risks at the micro and macro levels

i) **Competing risks at the micro-level.** We want to ensure that the two risks are effectively competing, namely that they have a comparable importance, even among advanced age survivors. In other words, for each individual, the probability that event 1 (resp. event 2) arrives first conditional on survival up to time t should not be vanishing :

$$\mathbb{P}[T_1 < T_2 | T_1 > t, T_2 > t, U, V] \nrightarrow 0, \text{or } 1.$$

The following property is proved in Appendix C.2 :

Property IV.3. Under the proportional hazard specification (IV-4), we have :

— If $\frac{\lambda_1(t)}{\lambda_2(t)} \to 0$, then $\mathbb{P}[T_1 < T_2 | T_1 > t, T_2 > t, U, V] \to 0$; for advanced age survivors, T_2 arrives nearly always first.

— If $\frac{\lambda_1(t)}{\lambda_2(t)} \to \infty$, then T_1 arrives nearly always first.

— If $\frac{\lambda_1(t)}{\lambda_2(t)} \to \ell$, then $\mathbb{P}[T_1 < T_2 | T_1 > t, T_2 > t, U, V] \to \frac{\ell U}{\ell U + V}$.

This property motivates the following assumption :

Assumption IV.2. We have $\lim_{t\to\infty} \dfrac{\lambda_1(t)}{\lambda_2(t)} = \ell > 0$, and $\lim_{t\to\infty} \Lambda_1(t) = \infty$.

As a consequence of Assumption IV.2, we get also :

$$\lim_{t\to\infty} \frac{\Lambda_1(t)}{\Lambda_2(t)} = \ell. \tag{IV-5}$$

This condition is a kind of asymptotic (deterministic) co-integration relationship. To highlight the importance of a common clock, we equivalently rewrite condition (IV-5) as :

$$\begin{cases} \Lambda_1(t) & \sim a_1 \Lambda(t) \\ \Lambda_2(t) & \sim a_2 \Lambda(t) \end{cases}, \text{ or } a_2\Lambda_1(t) - a_1\Lambda_2(t) = o(\Lambda(t)),$$

where $\Lambda(t) > 0$ and $a_1, a_2 > 0$ are positive constants such that $a_1/a_2 = \ell$.

ii) **Competing risks at the macro-level.** Property IV.3 focuses on micro level competition between the two risks. What happens when the unobserved heterogeneity (U, V) is integrated out ? Are they still effectively competing at the macro level, among advanced age survivors ? In other words, we want :

$$\mathbb{P}[T_1 < T_2 | T_1 > t, T_2 > t] \nrightarrow 0, \text{or } 1$$

to be satisfied. We have :

Property IV.4. Under the proportional hazard specification (IV-4) and Assumption IV.2, if the distribution of $\Lambda(t)(U, V)$ given $T_1 > t, T_2 > t$ converges to a limit distribution when t goes to infinity, then $\mathbb{P}[T_1 < T_2 | \min(T_1, T_2) = t]$ and $\mathbb{P}[T_1 < T_2 | \min(T_1, T_2) > t]$ converge to the same positive limit. In other words, the two risks are effectively, asymptotically competing at the macro level.

Proof : see Appendix C.2.

IV.3.2 Conditional distribution of (T_1, T_2) given $T_1 > t, T_2 > t$

As in the univariate case, by a change of variable we get an integral formula :

$$\mathbb{P}[T_1 > t + \tau_1, T_2 > t + \tau_2 | T_1 > t, T_2 > t] = \iint e^{-\left[\frac{\Lambda_1(t+\tau_1)}{\Lambda_1(t)} - 1\right] u^* - \left[\frac{\Lambda_2(t+\tau_2)}{\Lambda_2(t)} - 1\right] v^*} \mathrm{d}F_t^*(u^*, v^*),$$

where F_t^* is the cdf of $\Big(\Lambda_1(t)U, \Lambda_2(t)V\Big)$ given $T_1 > t, T_2 > t$. Thus let us consider the time change :

$$(X_t, Y_t) = \Big(\frac{\Lambda_1(T_1)}{\Lambda_1(t)} - 1, \frac{\Lambda_2(T_2)}{\Lambda_2(t)} - 1\Big).$$

When $T_1 > t, T_2 > t$ and t is large, we have, $\frac{\Lambda_1(T_j)}{\Lambda_1(t)} \sim \frac{\Lambda(T_j)}{\Lambda(t)}$, $j = 1, 2$. In other words, asymptotically, Λ is the common time change [4] for T_1, T_2, and $\min(T_1, T_2)$. Then we get :

Property IV.5. Under Assumption IV.2, if the cdf F_t of $\Lambda(t)(U, V)$ given $T_1 > t, T_2 > t$ converges [5] to F_∞, say, then the survivor function of (X_t, Y_t) given $T_1 > t, T_2 > t$ converges to the survivor function :

$$H(x, y) := \iint e^{-xu - yv} \mathrm{d}F_\infty(u/a_1, v/a_2), \quad x, y \geq 0. \tag{IV-6}$$

In other words, H is the survivor function of a bivariate survival variable with proportional hazard representation with unitary hazards and heterogeneity distribution $\mathrm{d}F_\infty(\frac{\cdot}{a_1}, \frac{\cdot}{a_2})$.

IV.3.3 Bivariate regular variation

Motivated by Properties IV.4 and IV.5, let us now look for a condition of convergence of the distribution $\Lambda(t)(U, V)$ given $T_1 > t, T_2 > t$.

Definition IV.1. We say that the cdf of (U, V) is regularly varying at $(0, 0)$, or $F \in BRV_0(\nu)$, if there exists a positive function ν such that for all $x, y > 0$:

$$\lim_{a \to 0} \frac{F(ax, ay)}{F(a, a)} = \lim_{a \to 0} \frac{\mathbb{P}[U < ax, V < ay]}{\mathbb{P}[U < a, V < a]} = \nu(x, y). \tag{IV-7}$$

4. This is important since we have (informally) $\mathbb{1}_{X_t < Y_t} \approx \mathbb{1}_{T_1 < T_2}$, namely, the order is (asymptotically) preserved under the time change.

5. Or, equivalently the cdf F_t^* of $\Big(\Lambda_1(t)U, \Lambda_2(t)V\Big)$ given $T_1 > t, T_2 > t$ converges to $F_\infty(u/a_1, v/a_2)$.

Since ν is non decreasing in both arguments and $\nu(0,0) = 0$, ν is a positive measure on $[0,\infty]^2$.

Theorem IV.2. Under Assumption 2 and mild regularity conditions, the distribution of $\Lambda(t)(U,V)$ given $T_1 > t, T_2 > t$ converges to a non degenerate distribution, if and only if the cdf of (U,V) is $BRV_0(\nu)$. In this case, the limit distribution is of the form :

$$\mathrm{d}F_\infty(u,v) = \frac{1}{c} e^{-a_1 u - a_2 v} \mathrm{d}\nu(u,v), \tag{IV-8}$$

where the normalizing constant $c = \iint e^{-a_1 u - a_2 v} \mathrm{d}\nu(u,v)$.

Proof : see Appendix C.2.

Which distributions are bivariate regularly varying at zero ? What are the corresponding measures ν ? Let us first provide a set of necessary conditions that the measure ν should satisfy, and then show that these conditions are sufficient for a function ν to qualify as a possible limit measure.

Property IV.6 (Necessary condition)**.** Measure ν is homogeneous of order $\alpha \geq 0$, namely, there exists α such that $\nu(cx,cy) = c^\alpha \nu(x,y), \quad \forall c,x,y \geq 0$.

Proof :

We have : $\frac{F(acx,acy)}{F(a,a)} = \frac{F(acx,acy)}{F(ac,ac)} \frac{F(ac,ac)}{F(a,a)}$. When a goes to zero, we get :

$$\nu(cx,cy) = \nu(x,y)\nu(c,c). \tag{IV-9}$$

Thus it suffices to prove that $\nu(c,c) = c^\alpha$. Let us set $x = y$ in equation (IV-9), we get $\nu(cx,cx) = \nu(x,x)\nu(c,c)$, for all x and c. Since the function $x \mapsto \nu(x,x)$ is positive and increasing, we easily deduce that $\nu(c,c) = c^\alpha, \quad \forall c > 0$, for some $\alpha \geq 0$. □

Measure ν can be singular, namely non absolutely continuous with respect to the Lebesgue measure on $[0,\infty[^2$. For instance, in the shared frailty model $U = V$, if the marginal cdf F_1 is $RV_0(\alpha)$, then :

$$\nu(x,y) = \lim_{a\to 0} \frac{F(ax,ay)}{F(a,a)} = \lim_{a\to 0} \frac{F_1(a\min[x,y])}{F_1(a)} = \min[x,y]^\alpha.$$

In this case, the limit heterogeneity distribution is concentrated on the diagonal $\{(x,x), x > 0\}$ and is marginally gamma distributed by the result of Section 2. This property can be equivalently

written in terms of the corresponding survivor copula [6] of (T_1, T_2).

For the rest of the paper, let us focus on non singular measures ν :

Assumption IV.3. Measure ν admits a positive Radon-Nikodym derivative μ with respect to the Lebesgue measure on $]0, \infty[^2$.

This density μ is also homogeneous, of order $\alpha - 2$, namely $\mu(cx, cy) = c^{\alpha-2}\mu(x, y)$ for all $c, x, y > 0$.

For each $\alpha > 0$, let us denote by $\mathcal{F}_\alpha$ the set of non singular measures ν that are homogeneous of order α and such that $\nu(1, 1) = 1$. This family is characterized by the following property :

Property IV.7. For any nonnegative function μ that is homogeneous of order $\alpha - 2$ and integrable on $[0, 1] \times [0, 1]$, its normalized primitive $\nu(x, y) := \frac{1}{c}\int_0^x \int_0^y \mu(u, v)\mathrm{d}u\mathrm{d}v$ belongs to $\mathcal{F}_\alpha$, where the normalizing constant is $c = \iint_0^1 \mu(u, v)\mathrm{d}u\mathrm{d}v$.

Remark 1. Because of the homogeneity, we can also introduce the spherical representation for μ, namely $\mu(r, \omega) = r^{\alpha-2}s(\omega)$, where $r = \sqrt{x^2 + y^2}$, and ω is defined by $x = r\cos\omega, y = r\sin\omega$. Then we remark that the integrability of μ on $[0, 1]^2$ is equivalent to its integrability on the disk $B = \{u^2 + v^2 \leq 1, u, v \geq 0\}$, which is equivalent to :

$$\iint_B \mu(u, v)\mathrm{d}u\mathrm{d}v = \int_0^{\frac{\pi}{2}} \int_0^1 r^{\alpha-2}s(\omega)r\mathrm{d}r\mathrm{d}\omega \propto \int_0^{\frac{\pi}{2}} s(\omega)\mathrm{d}\omega < \infty,$$

thus μ is integrable on $[0, 1]^2$ if and only if $\int_0^{\frac{\pi}{2}} s(\omega)\mathrm{d}\omega < \infty$.

Thus the family $\mathcal{F}_\alpha$ is large and semi-parametric. Can any $\nu \in \mathcal{F}_\alpha$ be interpreted as the limit measure in equation (IV-7), for some joint distribution (U, V) ? The answer is affirmative :

Property IV.8 (Sufficient condition)**.** A continuous bivariate distribution is regularly varying with measure $\nu(du, dv) = \mu(u, v)\mathrm{d}u\mathrm{d}v \in \mathcal{F}_\alpha$, if and only if its density is :

$$f(u, v) \propto l(u, v)\mu(u, v), \tag{IV-10}$$

6. Indeed, it is widely known that in a bivariate proportional hazard survival model with a shared frailty, the survivor copula of the two variables is an Archimedean copula whose generator is the Laplace transform of the frailty [see Marshall and Olkin (1988)]. Juri and Wüthrich (2002) show that if the generator is regularly varying at infinity, then the survivor copula for the residual lifetime converges to the Clayton copula, that is the Archimedean copula with gamma underlying frailty.

where function l is slowly varying at zero, namely, for all $x, y > 0$,

$$\lim_{a\to 0} \frac{l(ax, ay)}{l(a, a)} = 1, \tag{IV-11}$$

and satisfies the integrability condition $\iint l(u, v)\mu(u, v)\mathrm{d}u\mathrm{d}v < \infty$.

Proof : see Appendix C.2.

Let us now give examples of bivariate regularly varying distributions.

Example IV.2. The limit distribution in equation (IV-8) : $f(u, v) \propto e^{-a_1 u - a_2 v}\mu(u, v)$, is regularly varying at zero. Indeed, we have $\lim_{t\to\infty} e^{-a_1 u/t - a_2 v/t} = 1$ for all $u, v > 0$ and, *a fortiori*, this exponential function is slowly varying at zero. We can also verify that this distribution is properly defined. Indeed, we have :

$$\begin{aligned}\iint e^{-a_1 u - a_2 v}\mu(u, v)\mathrm{d}u\mathrm{d}v &= \sum_{n=0}^{\infty} \Big\{ \iint_{n\le u+v\le n+1} e^{-a_1 u - a_2 v}\mu(u, v)\mathrm{d}u\mathrm{d}v \Big\} \\ \le \sum_{n=0}^{\infty} \Big\{ e^{-\min(a_1,a_2)n} \iint_{[0,n+1]^2} \mu(u, v)\mathrm{d}u\mathrm{d}v \Big\} &= \sum_{n=0}^{\infty} \Big\{ e^{-\min(a_1,a_2)n}(n + 1)^{\alpha} \Big\} < \infty. \end{aligned} \tag{IV-12}$$

Example IV.3. Let us consider the measure :

$$\nu(x, y) = \int_0^{\alpha} \pi(\beta)x^{\beta}y^{\alpha-\beta}\mathrm{d}\beta,$$

with $\pi(\beta) \ge 0$, $\int_0^{\alpha} \pi(\beta)\mathrm{d}\beta = 1$, and $\beta \in (0, \alpha)$ for integrability reasons. Then the distribution $f(u, v) = e^{-a_1 u - a_2 v}\mu(x, y)$ is regularly varying, and is a mixture of gamma distributions.

Example IV.4. Model with correlated frailties [see Yashin et al. (1995)] is an extension of the shared frailty model. Let us set $U = G + F_1$ and $V = G + F_2$, where G, F_1, F_2 are independent variables. Assume that their distributions are RV_0 with indices δ, α, β, respectively. As in the univariate case, the distribution of :

$$\Big([\Lambda_1(t) + \Lambda_2(t)]G, \Lambda_1(t)F_1, \Lambda_2(t)F_2 \Big) \text{ given } T_1 > t, T_2 > t,$$

converges to a 3-dimensional gamma distribution with independent components : $\gamma(\delta, 1)$, $\gamma(\alpha, 1)$, and $\gamma(\beta, 1)$, respectively. Thus the limit distribution of $\Lambda(t)(U, V)$ given $T_1 > t, T_2 > t$ is a correlated gamma distribution. In this example, it is more convenient to consider the measure density μ instead of the measure ν. Indeed, we can compute the pdf of the limit distribution

by convolution of gamma distributions, and on the other hand, we know that this pdf is also $f_\infty(x,y) \propto e^{-a_1 x - a_2 y}\mu(x,y)$. Thus by identification, the density μ is necessarily of the form :

$$\mu(x,y) \propto \int_0^{\min(x,y)} u^{\delta-1}(x-u)^{\alpha-1}(y-u)^{\beta-1}\mathrm{d}u.$$

Example IV.5. If $U \in RV_0(\alpha_1)$, $V \in RV_0(\alpha_2)$ and (U,V) has a Gaussian copula with correlation coefficient $\rho \neq -1$, then (U,V) is regularly varying at zero, with $\nu(x,y) = x^{\frac{\alpha_1}{\rho+1}} y^{\frac{\alpha_2}{\rho+1}}$ [see Ledford and Tawn (1996)].

Example IV.6. If the heterogeneity (U,V) has marginal $\gamma(\alpha,1)$ distributions, with an Archimedean copula $C(x,y) = \psi^{-1}(\psi(x)+\psi(y))$, such that $\lim_{t\to\infty} \frac{\psi(at)}{\psi(t)} = a^{-\beta}$, then (U,V) is regularly varying at zero, with [see Charpentier and Segers (2009)] :

$$\nu(x,y) = \Big(\frac{x^{-\beta\alpha}}{2} + \frac{y^{-\beta\alpha}}{2}\Big)^{-\frac{1}{\beta}}.$$

IV.3.4 Properties of the new family of distributions

Theorem IV.2 provides a new semi-parametric family of distributions :

$$f(u,v) \propto e^{-a_1 u - a_2 v}\mu(u,v), \tag{IV-13}$$

where μ is homogeneous of order $\alpha - 2$, with $\alpha \geq 0$. Let us now study the properties of these distributions.

Property IV.9. Let us denote by (U,V) a couple following the distribution (IV-13), then :

i) If $\mu(x,y) \propto x^{\beta_1} y^{\beta_2}$ with $\beta_1, \beta_2 > -1$, then U, V are independent and gamma distributed.

ii) If $\alpha > 1$, then for any $\delta > 0$, the distribution of U conditional on $V/U = \delta$ is gamma : $\gamma(\alpha - 1, a_1 + a_2\delta)$.

iii) $a_1 U + a_2 V$ follows a gamma distribution $\gamma(\alpha, 1)$.

Proof : Property 9.1) is easily checked. For 9.2), it suffices to remark that, if $\alpha > 1$, then the density of U given $V/U = \delta$ exists and is such that :

$$f(u \mid v/u = \delta) \propto u^{\alpha-2} e^{-(a_1+a_2\delta)u},$$

which is the density of a gamma distribution $\gamma(\alpha - 1, a_1 + a_2\delta)$.

For 9.3), we have $\mathbb{E}[e^{-x(a_1U+a_2V)}] \propto \int e^{-a_1u(1+x)-a_2v(1+x)}\mu(u,v)\mathrm{d}u\mathrm{d}v = \frac{1}{(1+x)^\alpha}$ by homogeneity ; thus $a_1U + a_2V$ follows a gamma distribution $\gamma(\alpha, 1)$. □

Property 9.1) states that, if the initial distribution of heterogeneity is $BRV_0(\nu)$ with $\mu(x,y) \propto x^{\beta_1}y^{\beta_2}$ (see Example IV.5), then conditional on $T_1 > t, T_2 > t$, the normalized residual lifetimes (X_t, Y_t) are asymptotically independent for large t. Therefore, prior specifications on the distribution of heterogeneity (U, V) should be avoided since they may lead to degenerate asymptotics. Instead, the family of distributions $f(u,v) \propto e^{-a_1u-a_2v}\mu(u,v)$ provides a more flexible way of constructing regularly varying distributions (see Example IV.3).

As a consequence of Property 9.2), the distribution of (U, V) is a mixture of singular gamma components $(U, V = \delta U)$, for different values of δ. The mixing probability, that is the pdf h^* of $\delta = V/U$, is :

$$h^*(\delta) \propto \int uf(u,\delta u)\mathrm{d}u \propto \int e^{-(a_1+a_2\delta)u}u^{\alpha-1}\mu(1,\delta)\mathrm{d}u \propto \frac{\mu(1,\delta)}{(a_1+a_2\delta)^\alpha}. \tag{IV-14}$$

Let us now study how the heterogeneity distribution among survivors evolves, when the initial distribution follows $f(u,v) \propto e^{-a_1u-a_2v}\mu(u,v)$. For $t > 0$, the pdf of the heterogeneity among survivors is : $f_t(u,v) \propto e^{-[\Lambda_1(t)+a_1]u-[\Lambda_2(t)+a_2]v}\mu(u,v)$. Thus the distribution of U given $V/U = \delta, T_1 > t, T_2 > t$ is the gamma distribution $\gamma(\alpha - 1, a_1 + \Lambda_1(t) + a_2\delta + \Lambda_2(t)\delta)$, and the pdf of V/U among survivors is :

$$h^*(\delta|T_1 > t, T_2 > t) \propto \frac{\mu(1,\delta)}{(a_1 + \Lambda_1(t) + a_2\delta + \Lambda_2(t)\delta)^\alpha}.$$

Let us now interpret Property 9.3). Assume that the initial heterogeneity distribution of (U, V) is regularly varying at zero and consider the survival variable $T = \min(T_1, T_2)$. Its conditional intensity is :

$$\theta(t|T > t, U, V) = \lambda_1(t)U + \lambda_2(t)V \sim \lambda(t)(a_1U + a_2V).$$

Thus asymptotically, for large t, the remaining lifetime of T given $T > t$ satisfies a proportional hazard specification with heterogeneity $a_1U + a_2V$, and we shall expect similar results as in the univariate case (see Theorem 1). Indeed we have the following property :

Property IV.10. 1. If (U, V) is BRV_0, then $a_1U + a_2V$ is $RV_0(\alpha)$, for all $a_1, a_2 > 0$.

2. If the initial heterogeneity distribution (U, V) is BRV_0, and if the cumulated baseline hazards Λ_1, Λ_2 satisfy the co-integration Assumption 2, then the distribution of $\Lambda(t)(a_1U + a_2V)$ given $T_1 > t, T_2 > t$ converges to the gamma distribution $\gamma(\alpha, 1)$.

Proof : see Appendix C.2.

IV.3.5 Illustration

Asymptotic heterogeneity distribution. To illustrate the flexibility of the family of distributions $f(u, v) \propto e^{-a_1u-a_2v}\mu(u, v)$, with homogeneous measure density μ, in particular, its capability of generating either positive, or negative dependence between the two heterogeneity components, let us plot the iso-density curves for two limiting heterogeneity distributions, that are :

$$f_\infty(u, v) \propto e^{-0.5u-0.31v}\frac{u^2v^2}{u+v}, \qquad \text{in Figure 2,}$$
$$f_\infty(u, v) \propto e^{-0.4u-0.31v}(u^3 + v^3), \qquad \text{in Figure 3.}$$

In the first case, variables U and V are positively correlated and the correlation coefficient computed numerically is : $\text{Corr}(U, V) = 0.06$. They are negatively correlated : $\text{Corr}(U, V) = -0.49$ in the second case.

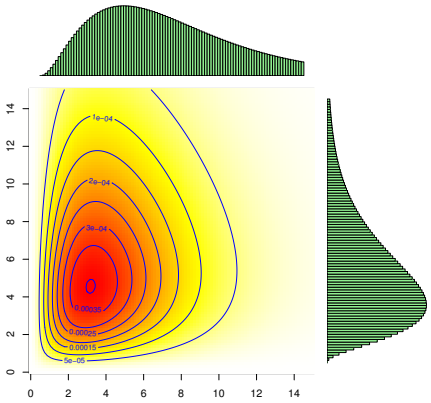


FIGURE IV-2: Iso-densities of $f_\infty(u, v) \propto e^{-0.5u-0.31v}\frac{u^2v^2}{u+v}$. Variables U and V are positively correlated.

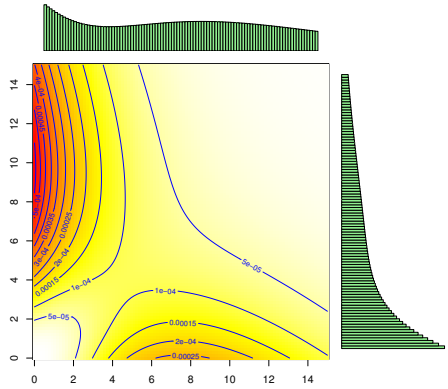


FIGURE IV-3: Iso-densities of $f_\infty(u,v) \propto e^{-0.4u-0.31v}(u^3+v^3)$. Variables U and V are negatively correlated.

Let us now comment on Figures 2 and 3. In each figure, the central panel displays the iso-densities. The first pdf has a unique global maximum at $(u,v) = (3.2, 4.5)$. In the second case, there are a unique global minimum at $(3.9, 4.5)$, as well as two local maxima on the boundary, respectively at $(0, 9.7)$ and $(7.5, 0)$. The two marginal pdf's are plotted on the upper and right panels, respectively.

Evolution of the heterogeneity distribution. Let us now illustrate the convergence of the heterogeneity distributions among the survivors to the previous limit distributions, when the initial distribution is regularly varying at zero. Let us define the initial heterogeneity distribution by :

$$f(u,v) = \frac{1}{3}\big(f_A(u,v) + f_B(u,v) + f_C(u,v)\big), \tag{IV-15}$$

where

$$f_A(u,v) \propto e^{-u-v}\frac{u^2v^2}{u+v}, \quad f_B(u,v) \propto e^{-u-v}uv^4, \quad f_C(u,v) \propto e^{-u-v}u^4v. \tag{IV-16}$$

This mixture distribution is jointly regularly varying at zero, with the same limit measure as the distribution f_A. Indeed, for fixed x, y, by the definition of bivariate regular variation, the univariate function $a \mapsto F_j(ax, ay)$ is univariate regularly varying at infinity, for $j \in \{A, B, C\}$, with index $-5, -7, -7$, respectively. Thus

$$F(ax,ay) = \frac{1}{3}\Big[F_A(ax,ay) + F_B(ax,ay) + F_C(ax,ay)\Big] \sim \frac{1}{3}F_A(ax,ay),$$

when a goes to zero, since the other two terms are negligible. In particular, $F(a,a) \sim \frac{1}{3}F_A(a,a)$, and

$$\lim_{a\to 0} \frac{F(ax,ay)}{F(a,a)} = \lim_{a\to 0} \frac{F_A(ax,ay)}{F_A(a,a)} = \nu_A(x,y),$$

where the measure ν_A has the density $\mu_A(x,y) \propto \frac{x^2y^2}{x+y}$.

Let us set the two cumulated baseline hazard functions as $\Lambda_1(t) = 0.5t, \quad \Lambda_2(t) = 0.31t$, for all $t > 0$. Thus we can take $\Lambda(t) = t$ and by Theorem 2, the limit distribution of $t(U,V)$ given $T_1 > t, T_2 > t$ is : $f_\infty(u,v) \propto e^{-0.5u-0.31v}\frac{u^2v^2}{u+v}$. We display, in Figure 4, the iso-densities of $\Lambda(t)(U,V)$ given $T_1 > t, T_2 > t$ for four different values of $t \in \{0, 0.3, 2, 5\}$.

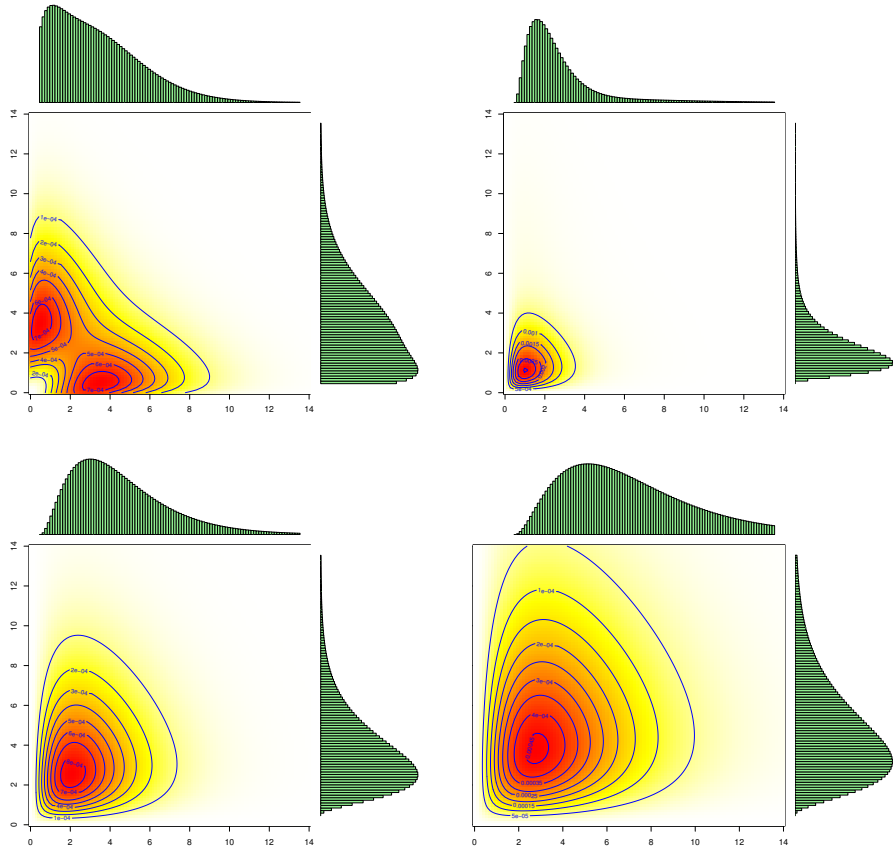


FIGURE IV-4: Evolution of the density of $\Lambda(t)(U,V)$ given $T_1 > t, T_2 > t$. Upper left panel : $t = 0$; upper right panel : $t = 1$, lower left panel : $t = 2$, lower right panel : $t = 5$.

In each of the four sub-figures, the iso-densities are given in the central panel plots, and the marginal pdf are given in the upper and right panels, respectively. The iso-densities evolve continuously when t increases, and for large t (that is $t = 5$), we get a distribution that is close to the limit distribution (see Figure 2). The marginal heterogeneity distributions among the

survivors also converge to the corresponding component of the limit distribution.

Let us now conduct the same analysis by changing the distribution f_A into :

$$f_A(u,v) \propto e^{-u-v}(u^3+v^3).$$

Let us set the cumulative hazard functions as $\Lambda_1(t) = 0.4t, \quad \Lambda_2(t) = 0.31t$, for all $t > 0$. Then the limit distribution of $t(U,V)$ given $T_1 > t, T_2 > t$ is $f_\infty(u,v) \propto e^{-0.4u-0.31v}(u^3+v^3)$, whose iso-densities are plotted in Figure 3. We display, in Figure 5, the iso-densities of $\Lambda(t)(U,V)$ given $T_1 > t, T_2 > t$ for four different values of $t \in \{0, 0.3, 2, 5\}$.

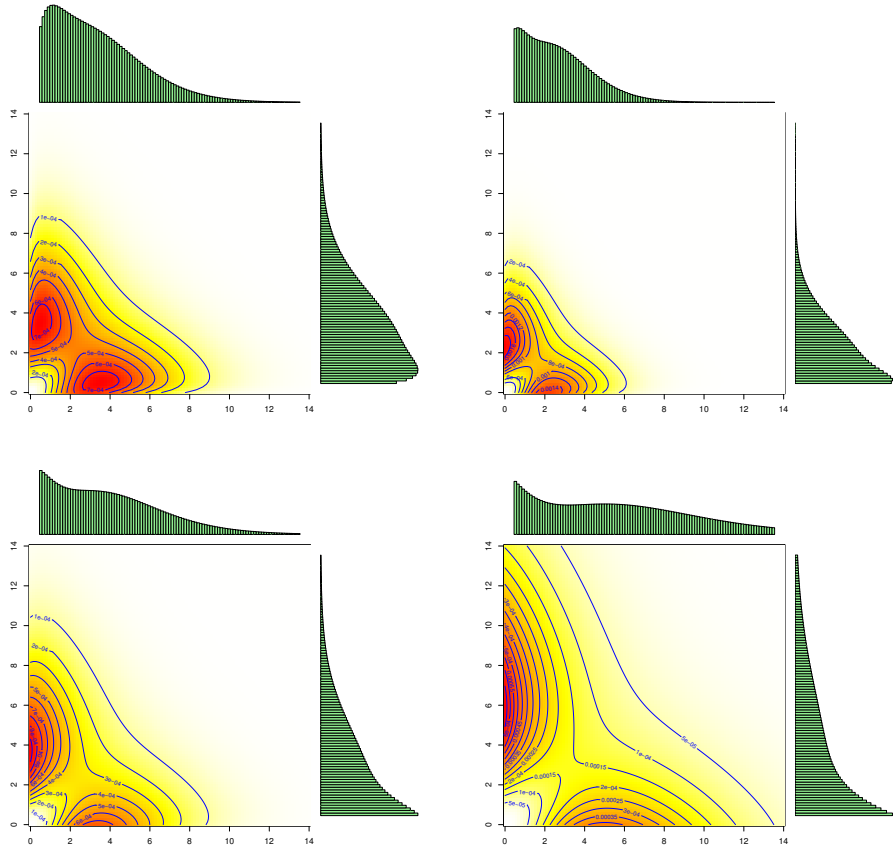


FIGURE IV-5: Evolution of the density of $\Lambda(t)(U,V)$ given $T_1 > t, T_2 > t$. Upper left panel : $t = 0$; upper right panel : $t = 1$, lower left panel : $t = 2$, lower right panel : $t = 5$.

IV.3.6 Marginal tails

i) **Tail of** $\min(T_1, T_2)$**.** As in the univariate case, we are interested in the survival probability $\mathbb{P}[T_1 > t, T_2 > t]$. We have the following property :

Property IV.11. Under the assumption of regular variation at zero of the heterogeneity distribution, and the co-integration Assumption 2, we have :

$$S(t,t) := \mathbb{P}[T_1 > t, T_2 > t] = \frac{L(\Lambda(t))}{\Lambda(t)^\alpha}, \tag{IV-17}$$

where L is slowly varying at infinity. Moreover, the hazard function h of $\min(T_1, T_2)$ is such that :

$$h(t) := -\frac{\mathrm{d}}{\mathrm{d}t} \log S(t,t) \sim \alpha \frac{\lambda(t)}{\Lambda(t)}. \tag{IV-18}$$

Proof : see Appendix C.2.

Thus, as in the univariate case (Section 2.3), the asymptotic behavior of the survival probability $\mathbb{P}[T_1 > t, T_2 > t]$ depends on $\Lambda(t)$ and α (up to a slowly varying function).

ii) **Marginal tails of T_1 and T_2.** Let us now show that generically, the bivariate regular variation assumption, as well as the co-integration assumption, do not constrain the marginal tail properties of T_1 or T_2. By Property IV.2, it suffices to show that these assumptions do not constrain the marginal left tail behavior of U and V. Indeed we have the following property :

Property IV.12. Under mild regularity conditions on $\nu \in \mathcal{F}_\alpha$, and for any indices α_1, α_2 no larger than α, there exists a couple (U, V) such that :

— the joint distribution of (U, V) is regularly varying with limit measure ν,

— the marginal distributions of U and V are regularly varying at zero, with indices α_1, α_2 respectively.

Proof : see Appendix C.2.

Note that the condition $\alpha_1 \leq \alpha$ (resp. $\alpha_2 \leq \alpha$) is sharp. Indeed, since $\mathbb{P}[U < a] \geq \mathbb{P}[U < a, V < a]$, if both the marginal distribution U and the joint distribution of (U, V) are regularly varying, then we have necessarily $\alpha_1 \leq \alpha$.

IV.4 Identification of parameters from advanced age survivors

Let us consider the identification of asymptotic parameters, which include the density μ, the scalars a_1, a_2, as well as the asymptotic behavior of $\Lambda(t)$. As seen in Property IV.5, these

parameters characterize the limit distribution of (T_1, T_2) given $T_1 > t, T_2 > t$. Can they be recovered from the data, under the co-integration and regular variation assumptions ?

Identification results have already been derived for mixed proportional hazard (MPH) models [see Honoré (1993); Abbring and van den Berg (2003a)], but our analysis is different : it concerns the asymptotic parameters and rely on observations of advanced age survivors only.

Without loss of generality, we can, by scale normalization, assume that $\ell = \lim_{t\to\infty} \frac{\Lambda_2(t)}{\Lambda_1(t)} = 1$ and $a_1 = a_2 = 1$. Indeed, we can always replace $(U, V, \Lambda_1, \Lambda_2)$ by $(a_1 U, a_2 V, \Lambda_1/a_1, \Lambda_2/a_2)$ without modifying the distribution of the duration variables.

In the following subsections, we will prove that, based on advanced age survivors, index α is never identified, $\Lambda(t)$ is always identified, while the identification of μ depends on the observability condition on (T_1, T_2).

IV.4.1 Non identification of α

Let us first show that the parameter α cannot be recovered from advanced age survivors only.

Property IV.13. It is not possible to identify the value of α if we observe only advanced age survivors.

Proof : The following two lemmas highlight observationally equivalent models.

Lemma IV.1. If $(\Lambda_1(t), \Lambda_2(t))$ and the distribution of (U, V) defines a proportional heterogeneity survival model, then, for all $\beta_1 > 1, \beta_2 > 1$, an observationally equivalent is obtained by $(\Lambda_1^{\beta_1}, \Lambda_2^{\beta_2})$ and the heterogeneity distribution of $(U_{\beta_1,\beta_2}, V_{\beta_1,\beta_2})$, whose Laplace transform is given by :

$$\mathcal{L}_{(U_{\beta_1,\beta_2}, V_{\beta_1,\beta_2})}(x, y) = \mathcal{L}_{(U,V)}(x^{1/\beta_1}, y^{1/\beta_2}), \qquad \forall x, y \geq 0. \tag{IV-19}$$

Proof : see Appendix C.3.

Moreover, this transformation does not compromise the regular variation property of the heterogeneity distribution. Indeed we have :

Lemma IV.2. If $\beta_1 = \beta_2$ and $(U, V) \in BRV_0(\nu)$, then the distribution of $(U_{\beta_1,\beta_2}, V_{\beta_1,\beta_2})$ is still $BRV_0(\nu_{\beta_1})$ with a limit measure ν_{β_1}, which is homogeneous of order α/β_1.

Proof : see Appendix C.3.

Therefore, we can always define an observationally equivalent model in which the heterogeneity distribution is regularly varying at zero, with an order α as small as we want. Thus α is not identifiable from advanced age survivors only. □

As a consequence of this property, we have the following corollary :

Property IV.14. By defining appropriately the heterogeneity distribution, we can fix the value of α as $\alpha = \alpha_0$, where α_0 is sufficiently small.

The observational equivalence in Lemma IV.1 extends a similar result derived in the univariate case by Ridder (1990); Ishwaran (1996). Since the derivative of the RHS of equation (IV-19) is infinite at $(0, 0)$, the transformed heterogeneities have necessarily infinite mean [7] : $\mathbb{E}[U_{\beta_1,\beta_2}] = \mathbb{E}[V_{\beta_1,\beta_2}] = \infty$.

To illustrate the infinite mean of the observationally equivalent heterogeneity distributions, let us consider an initial heterogeneity distribution with density $f(u, v) \propto e^{-u-v} \frac{u^2 v^2}{u+v}$. The observationally equivalent distribution is defined via its Laplace transform and it is difficult to obtain the closed form expression of the associated density function. Nevertheless, we display in Figure 6, the initial and transformed Laplace transforms (for $\beta = 1/3$), and x and y ranging from 0 to 4.

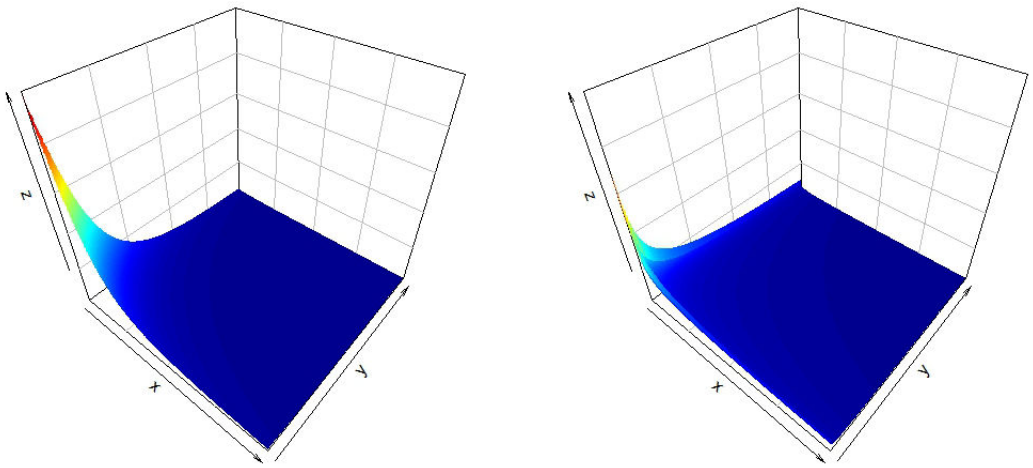


FIGURE IV-6: Left panel : the initial Laplace transform of $f(u, v) \propto e^{-u-v} \frac{u^2 v^2}{u+v}$. Right panel : the transformed Laplace transform.

7. This non identifiability result is compatible with the usual identifiability of MPH bivariate survival models which relies on the finite mean assumption [see e.g. Abbring and van den Berg (2003a)].

In both plots, the Laplace transform reaches its global maximum, which is equal to 1, at point $(x, y) = (0, 0)$. Its value is nearly zero for large values of x (or y), but this decay is faster for the transformed distribution. Indeed, for the initial distribution, the curve is differentiable at $x = 0$ and $y = 0$, with respect to y (resp. x) ; whereas for the transformed distribution, the partial derivatives at $x = 0$ or $y = 0$ are infinite.

IV.4.2 Identification of $\log \Lambda(t)$ for large t

Let us consider the information we can recover, if we only observe the minimum $T = \min(T_1, T_2)$ among advanced age survivors. From now on let us assume that we have fixed $\alpha = \alpha_0$. Then we have the following property :

Property IV.15. From the observation of $T = \min(T_1, T_2)$ among advanced age survivors, we can identify $\log \Lambda(t)$ for t large. Moreover we have :

$$\log \Lambda(t) \sim -\frac{1}{\alpha_0} \log S(t, t), \quad \text{for large } t. \tag{IV-20}$$

Proof : From the observation of T, we can identify $S(t, t)$. By Property IV.11, we have $S(t, t) = \frac{L(\Lambda(t))}{\Lambda^{\alpha_0}(t)}$. Therefore, given $\epsilon \in [0, 1[$, we have, for large t :

$$L(\Lambda(t))(1 - \epsilon) < S(t, t)\Lambda^{\alpha_0}(t) < L(\Lambda(t))(1 + \epsilon).$$

Since L is slowly varying at infinity, we have, for large t (see Lemma C.2 in Appendix C.1) :

$$\Lambda(t)^{-\epsilon} < L(\Lambda(t)) < \Lambda(t)^{\epsilon}.$$

By combining the previous two equations, we get, for large t :

$$-\epsilon \log(1 - \epsilon) < \frac{\log S(t, t)}{\log \Lambda(t)} + \alpha_0 < \epsilon \log(1 + \epsilon).$$

Thus $\log \Lambda(t) \sim \frac{1}{\alpha_0} \log S(t, t)$. □

IV.4.3 Identification of density μ

Let us finally discuss the identification of the functional parameter μ. We will see that the possibility to identify μ depends on the available observations.

The identification of $\log \Lambda(t)$ suggests to replace $\Lambda(t)$ by $S(t,t)^{-1/\alpha_0}$ in Property IV.5. The following property will be useful for the identification analysis of μ.

Lemma IV.3. Let us define $(\tilde{X}_t, \tilde{Y}_t)$ by :

$$(\tilde{X}_t, \tilde{Y}_t) := \Big(\frac{[S(t,t)]^{1/\alpha_0}}{[S(T_1,T_1)]^{1/\alpha_0}} - 1, \frac{[S(t,t)]^{1/\alpha_0}}{[S(T_1,T_1)]^{1/\alpha_0}} - 1 \Big), \tag{IV-21}$$

then the survivor function of $(\tilde{X}_t, \tilde{Y}_t)$ given $T_1 > t, T_2 > t$ converges to the survivor function :

$$H(x,y) = \frac{1}{c} \iint e^{-(1+x)u-(1+y)v} \mu(u,v) \mathrm{d}u\mathrm{d}v, \tag{IV-22}$$

where the constant $c = \iint e^{-u-v} uv\mu(u,v)\mathrm{d}u\mathrm{d}v$.

Proof : see Appendix C.3.

Observation of $(\min(T_1,T_2), \mathbb{1}_{T_1<T_2})$

Let us first consider the competing risks case, when we only observe $(\min(T_1,T_2), \mathbb{1}_{T_1<T_2})$. In this case the limiting survivor function H in equation (IV-22) cannot be identified and we have the following property :

Property IV.16. From the observation of $(\min(T_1,T_2), \mathbb{1}_{T_1<T_2})$ among advanced age survivors, we can not identify totally the density μ. The only identifiable functional of μ is :

$$\mathbb{P}[X < Y] = \frac{\iint e^{-u-v} u\mu(u,v)\mathrm{d}u\mathrm{d}v}{\alpha_0 \iint e^{-u-v}\mu(u,v)\mathrm{d}u\mathrm{d}v}.$$

Proof : Since $T_1 < T_2$ if and only if $\tilde{X}_t < \tilde{Y}_t$, we can only identify the distribution of couple $(\min[\tilde{X}_t, \tilde{Y}_t], \mathbb{1}_{\tilde{X}_t<\tilde{Y}_t})$. This distribution, conditional on $T_1 > t, T_2 > t$, converges to the (unconditional) distribution of :

$$(\min(X,Y), \mathbb{1}_{X<Y}), \tag{IV-23}$$

where the distribution of (X,Y) is defined by the survivor function (IV-22). Thus we can only identify, with the observations of advanced age survivors, the distribution of the couple (IV-23), which is characterized in the following lemma :

Lemma IV.4. The distribution of $\min(X, Y)$ is gamma $\gamma(\alpha_0, 1)$, and the conditional probability

$$p := \mathbb{P}[X < Y \mid \min(X, Y) = \tau] = \frac{\iint e^{-u-v} u\mu(u, v)\mathrm{d}u\mathrm{d}v}{\alpha_0 \iint e^{-u-v}\mu(u, v)\mathrm{d}u\mathrm{d}v},$$

which does not depend on τ.

Proof : see Appendix C.3.

As a consequence, only one functional of μ is identified, and it is not possible to identify totally the density μ. □

Remark 2. In other words, it is not possible to identify the distribution of (X, Y) from the distribution of $(\min(X, Y), \mathbb{1}_{X<Y})$. This result is analogous to the lack of identification for competing risks without covariates [see e.g. Tsiatis (1975)].

Observation of $(T_1 \mathbb{1}_{T_1<T_2}, T_2)$

Let us now consider the semi-competing risks, when only $(T_1 \mathbb{1}_{T_1<T_2}, T_2)$ is observed. We have the following property :

Property IV.17. The density μ is identifiable from the observation of $(T_1 \mathbb{1}_{T_1<T_2}, T_2)$.

Proof : We can identify the distribution of $(\tilde{X}_t \mathbb{1}_{\tilde{X}_t<\tilde{Y}_t}, \tilde{Y}_t)$ conditional on $T_1 > t, T_2 > t$, which converges to the distribution of $(X \mathbb{1}_{X<Y}, Y)$. This distribution has two components. The first component is on the domain $\mathcal{D}_1 = \{(x, y), 0 < x < y\}$, and the second one is degenerate and concentrated on the domain $\mathcal{D}_2 = \{(0, y), y > 0\}$. This distribution has a density with respect to $m_{\mathcal{D}_1} + m_{\mathcal{D}_2}$, where $m_{\mathcal{D}_j}$ is the Lebesgue measure on $\mathcal{D}_j$, $j = 1, 2$. This density is :

$$\begin{cases} h(x, y) &= \frac{1}{c} \iint e^{-(1+x)u-(1+y)v} uv\mu(u, v)\mathrm{d}u\mathrm{d}v \quad \text{on} \quad \mathcal{D}_1, \\ h(y) &= \frac{1}{c} \iint e^{-(1+y)u-(1+y)v} v\mu(u, v)\mathrm{d}v \qquad \text{on} \quad \mathcal{D}_2. \end{cases} \tag{IV-24}$$

The bivariate function $\iint e^{-(1+x)u-(1+y)v} uv\mu(u, v)\mathrm{d}u\mathrm{d}v$ is the Laplace-Stieltjes transform of the function $(u, v) \mapsto uv\mu(u, v)$, at point $(1 + x, 1 + y)$. Since this function is analytic, it is completely characterized by its values on domain $\mathcal{D}_1$. Therefore we can identify μ/c, by the uniqueness of this Laplace-Stieltjes transform. Since this constant is determined by the constraint $\iint_0^1 \mu(u, v)\mathrm{d}u\mathrm{d}v = 1$, μ is identified. □

IV.5 Conclusion

This paper considers bivariate survival models with bivariate proportional heterogeneity. We derive minimal conditions to ensure that the bivariate heterogeneity still exists among advanced age survivors. Then we show that, under further appropriate conditions on the heterogeneity distribution, the joint duration distribution among advanced age survivors admits a limit distribution.

Our model allows to capture the joint asymptotic dependence structure, without making constraints on the marginal distributions. Thus this approach contributes also to the extreme value theory for general multivariate variables. The current literature usually makes strong assumptions on the asymptotic behavior of marginal distributions [8] with a risk of mis-specification.

Finally, we get a new semi-parametric family of bivariate heterogeneity distributions, which arises as the limit heterogeneity distribution among advanced age survivors. This family is a serious competitor to the existing specifications of the bivariate heterogeneity distributions. Indeed, while the econometric literature traditionally emphasizes on the non-parametric identification of the unobserved heterogeneity, its implementation is very delicate. On the other hand, current parametric distributions are often too restrictive ; for instance, only the bivariate log-normal distribution has been proposed to capture negative dependence between different heterogeneity components [see Xue and Brookmeyer (1996)]. The new family offers a good trade-off between parsimony and flexibility : it is more parsimonious than an unconstrained distribution, and more flexible than current parametric heterogeneity distributions.

Appendix C.1 Univariate regular variation

This section provides proofs of the properties announced in Section 2, as well as technical lemmas on univariate regular variation.

Property C.1 (Alternative definition of regular variation)**.** A function F is RV_0 with index $\alpha > 0$ if and only if :

$$\lim_{y \to 0} \frac{F(y)}{y^{\alpha} L(y)} = 1,$$

where L is a slowly varying function at zero, that is, $\lim_{y \to 0} \frac{L(ay)}{L(y)} = 1$ for any $a > 0$.

8. It is often assumed that marginal distributions are asymptotically power law, with the same index.

Indeed, if $\frac{F(ay)}{F(y)}$ converges pointwise to a function $\lambda(a)$ of a, when y goes to zero, then $\lambda(a)$ is necessarily polynomial [see Feller (2008) (Lemma 1, VIII, 8)].

Definition C.1 (Regular variation at infinity)**.** A function H is regularly varying at infinity with index $\alpha \in \mathbb{R}$, or $H \in RV_\infty(\alpha)$, if for all $a > 0$, $\lim_{y\to\infty} \frac{H(ay)}{H(y)} = a^\alpha$, or equivalently if $H(y) = y^\alpha L(y)$, with L slowly varying at infinity : $\lim_{y\to\infty} \frac{L(ay)}{L(y)} = 1$.

Lemma C.1. [Inverse of a regularly varying function.] If a function $f > 0$ is increasing (resp. decreasing) and regularly varying at $+\infty$ with index $\alpha > 0$ (resp. $\alpha < 0$), then its inverse is also regularly varying at $+\infty$ (resp. 0) with index $1/\alpha$ (resp. $-1/\alpha$).

Proof : Assume, without loss of generality, that f is increasing and for any given $0 < x < 1$, we have : $\lim_{t\to\infty} \frac{f^{-1}(tx)}{f^{-1}(t)} = x^{1/\alpha}$. It suffices to show that for a given ϵ,

$$x^{1/\alpha+\epsilon} f^{-1}(t) \leq f^{-1}(tx) \leq x^{1/\alpha-\epsilon} f^{-1}(t), \tag{C-1}$$

for t large enough. Since

$$\lim_{t\to\infty} \frac{1}{t} f(x^{1/\alpha-\epsilon} f^{-1}(t)) = x^{1-\alpha\epsilon} > x = \frac{1}{t} f(f^{-1}(tx)),$$

the second inequality in (C-1) is satisfied for large t. The first one is derived similarly. □

Lemma C.2 (Limit of a regularly varying function, see e.g. Embrechts et al. (1997), Corollary A3.3)**.** Assume that function H is regularly varying at infinity with index $\alpha \in \mathbb{R}$,

— if $\alpha > 0$, $H(t)$ goes to infinity, when t goes to infinity.

— if $\alpha < 0$, $H(t)$ goes to 0, when t goes to infinity.

Therefore, slowly varying functions (at infinity) are negligible with respect to any positive power functions since $\frac{L(y)}{y^\alpha} \to 0$ when y goes to infinity, where L is slowly varying at infinity, for $\alpha > 0$.

Lemma C.3 (Uniform convergence of regular variation, see e.g. Embrechts et al. (1997) Theorem A3.2)**.** If f is regularly (or slowly) varying at infinity with index α, then for any $0 < a \leq b < \infty$, $\lim_{t\to\infty} \frac{f(tx)}{f(t)} \to x^\alpha$ uniformly in x,

— on each $[a, b]$ if $\alpha = 0$

— on each $(0, b]$ if $\alpha \neq 0$.

Theorem C.1 (See Feller (2008), XIII 5, Theorem 3)**.** Let F be the cdf of a positive variable U, then the following two properties are equivalent :

— The Laplace-Stieltjes transform $\mathcal{L}_F = \int e^{-ux} F(du)$ is $RV_\infty(-\alpha)$, $\alpha \geq 0$, namely, $\mathcal{L}_F(x) = \frac{1}{x^\alpha} L(x)$, where L is slowly varying at infinity.

— F is $RV_0(\alpha) : F(u) = u^\alpha L^*(u)$, where L^* is slowly varying.

Remark 3. One might expect that, if F is regularly varying with index α, then its derivative f is also regularly varying with index $\alpha - 1$, and that the previous equivalence theorem could be written in terms of f as well. This is true under some regularity conditions, for instance if f is monotone beyond a certain threshold [see Feller (2008) for a discussion. The same remark applies to bivariate cdf's [see de Haan and Resnick (1979); de Haan and Omey (1984) for technical conditions].

Appendix C.2 Bivariate regular variation

This appendix provides proofs of the properties announced in Section 3, as well as other useful properties regarding the notion of bivariate regular variation.

Lemma C.4 (No simultaneous arrival)**.** In a general bivariate survival model, if the joint density function is continuous, then the intensity that two events arrive simultaneously is zero :

$$\lim_{u \to 0} \frac{\mathbb{P}(T_1 \leq t+u, T_2 \leq t+u | \min(T_1, T_2) > t)}{u} = 0. \tag{C-2}$$

Proof : We have :

$$\mathbb{P}(T_1 \leq t+u, T_2 \leq t+u | \min(T_1, T_2) > t) = \frac{1}{S(t,t)} \Big(1 + S(t+u, t+u) - S(t+u, t) - S(t, t+u) \Big). \tag{C-3}$$

The limit, as well as the derivative of the RHS with respect to u, are equal to 0 at point $u = 0$. Thus we get equation (C-2). □

Corollary C.1 (Interpretation of the hazard functions.)**.** Under the proportional hazard specification (IV-4), we have :

1. $h(t) = h_1(t) + h_2(t)$, $\lambda(t|U,V) = \lambda_1(t)U + \lambda_2(t)V$, where h is the hazard function of $\min(T_1, T_2)$, and $\lambda(t|U,V)$ its conditional intensity given (U,V).

2. $h_1(t) = \lambda_1(t)\mathbb{E}[U|T_1 > t, T_2]$ and similarly for $h_2(t)$.

3. $\mathbb{P}[T_1 < T_2 | \min(T_1, T_2) = t] = \frac{h_1(t)}{h_1(t)+h_2(t)}$, $\mathbb{P}[T_1 < T_2 | \min(T_1, T_2) = t, U, V] = \frac{\lambda_1(t)U}{\lambda_1(t)U+\lambda_2(t)V}$.

Proof : Under the proportional hazard specification, both the unconditional and conditional densities of T_1, T_2 given U, V are continuous. Applying Lemma C.4, we get :

$$\begin{aligned} h_1(t) &= \lim_{\mathrm{d}t\to\infty} \frac{1}{\mathrm{d}t}\mathbb{P}[t < \min(T_1, T_2) < t + \mathrm{d}t, T_1 < T_2 \mid \min(T_1, T_2) > t] \\ &= \lim_{\mathrm{d}t\to\infty} \frac{1}{\mathrm{d}t}\mathbb{P}[t < T_1 < t + \mathrm{d}t, T_1 < T_2 \mid \min(T_1, T_2) > t] = -\frac{\partial}{\partial t_1} \log S(t, t), \end{aligned}$$

and similarly $h_2(t) = -\frac{\partial}{\partial t_2} \log S(t, t)$. Since $h(t) = -\frac{\mathrm{d}}{\mathrm{d}t} \log S(t, t)$ by definition, we get $h_1(t) + h_2(t) = h(t)$. The proof of the equality $\lambda_1(t)U + \lambda_2(t)V = \lambda(t|U, V)$ is similar by replacing all the unconditional probabilities/densities/intensities by their conditional counterparts.

For part 2), we have :

$$h_1(t) = -\frac{\partial}{\partial t_1} \log S(t, t) = \frac{\lambda_1(t)\mathbb{E}[Ue^{-\Lambda_1(t)U-\Lambda_2(t)V}]}{\mathbb{E}[e^{-\Lambda_1(t)U-\Lambda_2(t)V}]} = \lambda_1(t)\mathbb{E}[U|T_1 > t, T_2 > t].$$

Property 3) in Corollary A.1 is a direct consequence of Property 1. □

Proof of Property IV.3. We have :

$$\mathbb{P}[T_1 < T_2|T_1 > t, T_2 > t, U, V] = \frac{\int_t^\infty \mathbb{P}[T_1 < T_2 | \min(T_1, T_2) = \tau, U, V] f(\tau|U, V)\mathrm{d}\tau}{\mathbb{P}[T_1 > t, T_2 > t]},$$

where $f(\tau|U, V)$ is the conditional density of $\min(T_1, T_2)$ given U, V. By Corollary C.1,

$$\mathbb{P}[T_1 < T_2 | \min(T_1, T_2) = \tau, U, V] = \frac{\lambda_1(\tau)U}{\lambda_1(\tau)U + \lambda_2(\tau)V};$$

thus, if $\dfrac{\lambda_1(t)}{\lambda_2(t)}$ converges to 0, ∞, or ℓ, then

$$\mathbb{P}[T_1 < T_2 | \min(T_1, T_2) = \tau, U, V]$$

converges to 0, 1, and $\frac{\ell U}{\ell U+V}$, respectively. Thus we get the convergence of $\mathbb{P}[T_1 < T_2|T_1 > t, T_2 > t, U, V]$ to the same limit. □

Proof of Property IV.4. By Corollary C.1, we have :

$$\frac{h_1(t)}{h_2(t)} = \frac{\lambda_1(t)\mathbb{E}[U|T_1 > t, T_2 > t]}{\lambda_2(t)\mathbb{E}[V|T_1 > t, T_2 > t]} \sim \ell \frac{\mathbb{E}[\Lambda(t)U|T_1 > t, T_2 > t]}{\mathbb{E}[\Lambda(t)V|T_1 > t, T_2 > t]},$$

which converges to a positive number ℓ'. Thus $\mathbb{P}[T_1 < T_2 | \min(T_1, T_2) = t] = \frac{h_1(t)}{h_1(t)+h_2(t)}$ converges to $\frac{\ell'}{1+\ell'}$.

The convergence of $\mathbb{P}[T_1 < T_2 | \min(T_1, T_2) > t]$ can be proved in the same way as in Property IV.3 by replacing conditional probabilities/densities by their unconditional counterparts. □

Proof of Theorem IV.2. The convergence of the distribution of $(\Lambda(t)U, \Lambda(t)V)$ given $T_1 > t, T_2 > t$ is equivalent to the convergence of the distribution of $(\Lambda_1(t)U, \Lambda_2(t)V)$ given $T_1 > t, T_2 > t$, or to the convergence of the Laplace-Stieltjes transform of the cdf of the latter. This transform is equal to :

$$\begin{aligned}\mathcal{L}_2(x,y) &= \frac{\mathbb{E}\left[e^{-\Lambda_1(t)Ux-\Lambda_2(t)Vy}e^{-\Lambda_1(t)U-\Lambda_2(t)V}\right]}{\mathbb{E}\left[e^{-\Lambda_1(t)U-\Lambda_2(t)V}\right]} = \frac{\mathbb{E}\left[e^{-(x+1)\Lambda_1(t)U-(y+1)\Lambda_2(t)V}\right]}{\mathbb{E}\left[e^{-\Lambda_1(t)U-\Lambda_2(t)V}\right]} \\ &= \frac{\iint e^{-(1+x)u-(1+y)v} dF(\frac{u}{\Lambda_1(t)}, \frac{v}{\Lambda_2(t)})}{\mathbb{E}\left[e^{-\Lambda_1(t)U-\Lambda_2(t)V}\right]}\end{aligned}$$

Thus its convergence implies the pointwise convergence of : $\frac{F(\frac{u}{\Lambda_1(t)}, \frac{v}{\Lambda_2(t)})}{\mathbb{E}\left[e^{-\Lambda_1(t)U-\Lambda_2(t)V}\right]}$ to a measure $k(u,v)$, say. In particular, by taking $u = v = 1$ we also get : $\frac{F(\frac{u}{\Lambda_1(t)}, \frac{v}{\Lambda_2(t)})}{F(\frac{1}{\Lambda_1(t)}, \frac{1}{\Lambda_2(t)})} \to \frac{k(u,v)}{k(1,1)}$, or equivalently (by the monotonic property of F) : $\frac{F(\frac{u}{\Lambda(t)}, \frac{v}{\Lambda(t)})}{F(\frac{1}{\Lambda(t)}, \frac{1}{\Lambda(t)})} \to \frac{k(u/a_1, v/a_2)}{k(1/a_1, 1/a_2)}$. Thus F is regularly varying at zero.

Conversely, if F is regularly varying at zero, then, under some regularity conditions, the extended continuity theorem (Feller (2008), Theorem 2, Chapter XIII.1) applies and the previous steps can be reversed. □

Definition C.2. A bivariate function g is regularly varying at infinity (BRV_∞) if there exists $\phi > 0$ such that for all $x, y > 0$,

$$\lim_{t\to\infty} \frac{g\big(tx, ty\big)}{g\big(t, t\big)} = \phi(x, y).$$

Theorem IV.2 is called an Abel-Tauber theorem and we can similarly prove that :

Theorem C.2. [See de Haan et al. (1984)] A distribution F is $BRV_0(\nu)$ if and only if the Laplace-Stieltjes transform of F is regularly varying at infinity, that is, there exists a function $\phi > 0$ such that for all $x, y > 0$:

$$\lim_{t\to\infty} \frac{\mathcal{L}_F(tx, ty)}{\mathcal{L}_F(t, t)} = \phi(x, y).$$

In this case, $\phi = \frac{\mathcal{L}_\nu(\cdot,\cdot)}{\mathcal{L}_\nu(1,1)}$, where $\mathcal{L}_\nu(x, y) = \iint e^{-ux-vy} \mathrm{d}\nu(u, v)$ is the Laplace-Stieltjes transform of ν.

Proof of Property IV.8. If (U, V) follows (IV-10), then the pdf of $\Lambda(t)(U, V)$ given $T_1 > t, T_2 > t$ is

$$f_t(u, v) \propto e^{-\frac{\Lambda_1(t)}{\Lambda(t)}u - \frac{\Lambda_2(t)}{\Lambda(t)}v} l(\frac{u}{\Lambda(t)}, \frac{v}{\Lambda(t)})\mu(\frac{u}{\Lambda(t)}, \frac{v}{\Lambda(t)}). \tag{C-4}$$

Since $\Lambda_1(t) \sim a_1\Lambda(t)$, $\Lambda_2(t) \sim a_2\Lambda(t)$, $l(\frac{u}{\Lambda(t)}, \frac{v}{\Lambda(t)}) \sim l(\frac{1}{\Lambda(t)}, \frac{1}{\Lambda(t)})$, and μ is homogeneous, the distribution (C-4) converges to : $f_\infty(u, v) \propto e^{-a_1u-a_2v}\mu(u, v)$. Therefore, (U, V) is $BRV_0(\nu)$ by Theorem IV.2.

Conversely, if the cdf F is regularly varying at zero with a limit measure ν, then, under some regularity conditions [see Remark 3], the pdf f is also regularly varying, with :

$$\lim_{a\to 0} \frac{f(ax, ay)}{f(a, a)} = \frac{\mu(x, y)}{\mu(1, 1)}. \tag{C-5}$$

Let us define $l(x, y) = \frac{f(x,y)}{\mu(x,y)}$; then by equation (C-5), we can check that l is slowly varying at zero. □

Proof of Property IV.10. 1) By Theorem C.1, the regular variation at zero of $a_1U + a_2V$ is equivalent to the regular variation of its Laplace transform at infinity. When t goes to infinity, we have :

$$\frac{\mathbb{E}[e^{-(a_1U+a_2V)ct}]}{\mathbb{E}[e^{-(a_1U+a_2V)t}]} = \frac{\mathbb{E}[e^{-(a_1U+a_2V)ct}]/\mathbb{E}[e^{-(U+V)t}]}{\mathbb{E}[e^{-(a_1U+a_2V)t}]/\mathbb{E}[e^{-(U+V)t}]} \sim \frac{\mathcal{L}_\nu(a_1c, a_2c)}{\mathcal{L}_\nu(a_1, a_2)} = c^{-\alpha}$$

by the homogeneity of ν. Thus $a_1U + a_2V$ is $RV_0(\alpha)$.

Part 2) of Property IV.10 is a consequence of Property IV.9.1). □

Proof of Property IV.11. For a fixed $\epsilon > 0$, we have, for large t :

$$\Lambda_1(t) < (1+\epsilon)a_1\Lambda(t), \quad \Lambda_2(t) < (1+\epsilon)a_2\Lambda(t).$$

Thus :

$$\mathbb{E}\big[e^{-\Lambda_1(t)U-\Lambda_2(t)V}\big] \geq \mathbb{E}\big[e^{-(1+\epsilon)a_1\Lambda(t)U-(1+\epsilon)a_2\Lambda(t)V}\big] = \frac{L(\Lambda(t))}{\Lambda^\alpha(t)(1+\epsilon)^\alpha},$$

by the regular variation of $a_1U + a_2V$ and Theorem C.1. Similarly, we have :

$$\mathbb{E}\big[e^{-\Lambda_1(t)U-\Lambda_2(t)V}\big] \leq \frac{L(\Lambda(t))}{\Lambda^\alpha(t)(1-\epsilon)^\alpha}.$$

Hence the equivalence (IV-17).

Part 2) of Property IV.11 is a direct consequence of Property IV.10.2). □

Proof of Property IV.12. Let us first prove the following lemma :

Lemma C.5. If (U, V) follows $f(u, v) \propto e^{-a_1u-a_2v}\mu(u, v)$, then U (resp. V) is marginally regularly varying at zero if and only if the function $\delta \mapsto \mu(1, \delta)$ (resp. function $\delta \mapsto \mu(\delta, 1)$) is regularly varying at 0, with a nonnegative index. In both cases, the indices α_1 and α_2 are smaller than α.

Proof : The marginal pdf of U is :

$$f_1(u) \propto \int_0^\infty e^{-a_1u-a_2v}\mu(u, v)\mathrm{d}v = e^{-a_1u}u^{\alpha-1}\int_0^\infty e^{-a_2\delta u}\mu(1, \delta)\mathrm{d}\delta.$$

Under appropriate regularity conditions, the regular variation of the cdf is equivalent to that of the pdf, which is equivalent to that of the integral in the previous formula. This integral is the Laplace transform of function $\delta \mapsto \mu(1, \delta)$ at point a_2u. Then by an Abel-Tauber theorem [see Feller (2008), XIII 5, Theorem 1], this Laplace transform is regularly varying at zero with index $-\beta_1 \leq 0$ if and only if $\delta \mapsto \mu(1, \delta)$ is regularly varying at infinity with index $\beta_1 \geq 0$. In this case, the density f_1 is regularly varying at zero with index $\alpha - 1 - \beta_1$, and the cdf F_1 is regularly varying at zero with index $\alpha_1 = \alpha - \beta_1 \leq \alpha$ (see Remark 3). □

Now let us prove Property IV.12. For any $\alpha > 0$, consider the following distribution :

$$f(u,v) \propto e^{-u^{\alpha_3}-v^{\alpha_3}}\mu(u,v). \tag{C-6}$$

We can verify that :

1. This distribution is well defined. The proof mimicks inequality (IV-12).
2. This distribution is regularly varying with measure density μ. Indeed, it suffices to remark that $e^{-u^{\alpha_3}-v^{\alpha_3}}$ goes to 1 when (u,v) goes to zero.

Let us now compute the marginal pdf of this distribution. We have,

$$f_1(u) = e^{-u^{\alpha_3}}u^{\alpha-1}\int e^{-u^{\alpha_3}\delta^{\alpha_3}}\mu(1,\delta)\mathrm{d}\delta = \frac{1}{\alpha_3}e^{-u^{\alpha_3}}u^{\alpha-1}\int e^{-u^{\alpha_3}z}\mu(1,z^{1/\alpha_3})z^{1/\alpha_3-1}\mathrm{d}\delta.$$

The latter integral is the Laplace transform of the function $z \mapsto \mu(1,z^{1/\alpha_3})z^{1/\alpha_3-1}$ taken at argument u^{α_3}. Thus, by an Abel-Tauber theorem, it is regularly varying at zero with index $\alpha_3(\beta_1/\alpha_3+1/\alpha_3-1)$, so long as this index is nonnegative. Thus U (resp. V) is regularly varying with index $\alpha-(\beta_1+1-\alpha_3)$, so long as $\beta_1+1-\alpha_3$ (resp. $\beta_2+1-\alpha_3$) is nonnegative.

If $\beta_1-\beta_2 = \alpha_1-\alpha_2$, then we can take α_3 which satisfies simultaneously :

$$\alpha-(\beta_1+1-\alpha_3) = \alpha_1, \tag{C-7}$$

$$\alpha-(\beta_2+1-\alpha_3) = \alpha_2. \tag{C-8}$$

Thus we have already constructed a distribution by (C-6), which satisfies all the desired properties. From now on let us assume, without loss of generality, that $\beta_1-\beta_2 < \alpha_1-\alpha_2$. Let us define α_3 by equation (C-7). Then the regular variation index of V is $\alpha-(\beta_2+1-\alpha_3) > \alpha_2$.

To obtain a new distribution which is still jointly and marginally regularly varying, with respectively marginal regular variation indices α_1 and α_2, we consider, as in Section 3.5, the mixture distribution :

$$f_3(u,v) \propto f(u,v) + e^{-u-v}u^{\alpha+1}v^{\alpha_2},$$

where f is defined in equation (C-6). We can easily check that both the joint and marginal regular variation properties remain valid, except that the index of V is replaced by α_2. □

We end this appendix with some remarks on the previous proof.

Remark 4. We can similarly construct a distribution that is jointly, but not marginally regularly

varying. For instance, the distribution defined by :

$$f_4(u,v) \propto f(u,v) + u^{\alpha_1(1-\epsilon)}(2+\sin u)v^{\alpha+1} + v^{\alpha_2(1-\epsilon)}(2+\sin v)u^{\alpha+1},$$

is not marginally regularly varying, for any $\epsilon > 0$. In fact, the marginal distribution of U is asymptotically proportional to $u^{\alpha_1(1-\epsilon)}(2+\sin u)$.

Remark 5. Let us now show that when functions $\delta \mapsto \mu(1,\delta)$ and $\delta \mapsto \mu(\delta,1)$ are regular varying, their indices α_1 and α_2 are "nuisance parameters" ; in other words, these two regular variation conditions are quite mild conditions.

Let us consider the heterogeneity distribution plotted in Figure 5. In this example, we have $\mu(u,v) = \frac{u^2v^2}{u+v}$; thus $\alpha - 2 = 3$, $\beta_1 = 2, \beta_2 = 1$, $\alpha_1 = 3, \alpha_2 = 4$. It is easily checked that for any positive t, the marginal distribution of the heterogeneity U given $T_1 > t, T_2$ is regularly varying, with index 2. This index is equal to the regular variation index of the initial distribution. Therefore, the marginal tail behavior of U among survivors is always different from $\alpha_1 = 3$, that is the regular variation index of the limit distribution.

Thus although the marginal distributions converge, their regular variation indices at zero do not, in general. Therefore, the value of α_1 can never be correctly recovered from the knowledge at a finite date t.

Therefore, without loss of generality, we can assume, say, $\beta_1 = \beta_2 = 0$.

Appendix C.3 Identification

This section provides the proofs of properties announced in Section 4.

Proof of Lemma IV.1. It suffices to prove that equation (IV-19) defines a bivariate, positive distribution. We remark that the RHS of equation (IV-19) is a bivariate completely monotone function, that is,

$$(-1)^{n_1+n_2}\frac{\partial^{n_1+n_2}}{\partial x^{n_1}\partial y^{n_2}}\mathbb{E}[e^{-x^{1/\beta_1}U - y^{1/\beta_2}V}] \geq 0, \qquad \forall n_1, n_2 \in \mathbb{N}, x, y > 0.$$

Then, by a multivariate extension of the Bernstein-Widder theorem [see Berg et al. (1984), Zocher (2006)], it is the Laplace transform of a bivariate positive variable. □

Proof of Lemma IV.2. We have :

$$\lim_{t\to\infty}\frac{\mathcal{L}_{(U_{\beta_1,\beta_1},V_{\beta_1,\beta_1})}(tx,ty)}{\mathcal{L}_{(U_{\beta_1,\beta_1},V_{\beta_1,\beta_1})}(t,t)}=\lim_{t\to\infty}\frac{\mathcal{L}_{(U,V)}((tx)^{1/\beta_1},(ty)^{1/\beta_1})}{\mathcal{L}_{(U,V)}(t^{1/\beta_1},t^{1/\beta_1})}=\frac{\mathcal{L}_{\nu}(x^{1/\beta_1},y^{1/\beta_1})}{\mathcal{L}_{\nu}(1,1)}.$$

Thus by Theorem C.2, the distribution of $(U_{\beta_1,\beta_2},V_{\beta_1,\beta_2})$ is regularly varying with a limit measure ν_{β_1} such that $\frac{\mathcal{L}_{\nu}(x^{1/\beta_1},y^{1/\beta_1})}{\mathcal{L}_{\nu}(1,1)}=\frac{\mathcal{L}_{\nu_{\beta_1}}(x,y)}{\mathcal{L}_{\nu_{\beta_1}}(1,1)}$. Thus the homogeneity order of ν_{β_1} is α/β_1. □

Proof of Lemma IV.3. Denote by $H_t(x,y)$ the survivor function of this conditional distribution, and S^{-1} the inverse of $S(t,t)$, then we have :

$$\begin{aligned}
\lim_{t\to\infty} LHS &= \lim_{t\to\infty}\frac{\mathbb{P}[\frac{S(t,t)}{S(T_1,T_1)}>(1+x)^{\alpha},\frac{S(t,t)}{S(T_2,T_2)}>(1+y)^{\alpha}]}{\mathbb{P}[T_1>t,T_1>t]}\\
&= \lim_{t\to\infty}\frac{\mathbb{P}\big[\Lambda(T_1)>\Lambda\circ S^{-1}[\frac{S(t,t)}{(1+x)^{\alpha}}],\Lambda(T_2)>\Lambda\circ S^{-1}[\frac{S(t,t)}{(1+y)^{\alpha}}]\big]}{\mathbb{P}[T_1>t,T_1>t]} &\text{(C-9)}\\
&= \lim_{t\to\infty}\frac{\mathbb{P}\big[\Lambda(T_1)>(1+x)\Lambda(t),\Lambda(T_2)>(1+y)\Lambda(t)\big]}{\mathbb{P}[T_1>t,T_1>t]} &\text{(C-10)}\\
&= \lim_{t\to\infty}\frac{\mathcal{L}_{(U,V)}\left\{\Lambda_1\circ\Lambda^{-1}[(1+x)\Lambda(t)],\Lambda_2\circ\Lambda^{-1}[(1+y)\Lambda(t)]\right\}}{\mathcal{L}_{(U,V)}(\Lambda_1(t),\Lambda_2(t))} &\text{(C-11)}\\
&= \lim_{t\to\infty}\frac{\mathcal{L}_{(U,V)}\left\{(1+x)\Lambda_1(t),(1+y)\Lambda_2(t)\right\}}{\mathcal{L}_{(U,V)}(\Lambda_1(t),\Lambda_2(t))}=\frac{\mathcal{L}_{\mu}(1+x,1+y)}{\mathcal{L}_{\mu}(1,1)}=H(x,y). &\text{(C-12)}
\end{aligned}$$

To get equality (C-10), we have used the regular variation of $\Lambda\circ S^{-1}$. Indeed, $S(t,t)=\frac{L(\Lambda(t))}{\Lambda^{\alpha}(t)}$; thus $S\circ\Lambda^{-1}$ is $RV_{\infty}(-\alpha)$, and its inverse is $RV_0(-\frac{1}{\alpha})$ (see Lemma C.1). We can replace the different quantities by equivalent ones because of the monotonicity property of the survivor function and the continuity of the Laplace transform $\mathcal{L}_{\mu}$. To get equality (C-12), we have used the fact that $\Lambda_1\sim a_1\Lambda$ and $\Lambda_2\sim a_2\Lambda$. □

Remark 6. Property IV.3 is a generalization of the property of regular variation at infinity. Indeed, if Λ is regularly varying at infinity : $\lim_{t\to\infty}\frac{\Lambda(at)}{\Lambda(t)}=a^{\beta}$, for all $a>0$, where $\beta>0$, then we have $\mathbb{P}[\min(T_1,T_2)>t]=\frac{L(\Lambda(t))}{\Lambda^{\alpha}(t)}=\frac{L_2(t)}{t^{\beta\alpha}}$, and $\min(T_1,T_2)$ is heavy-tailed. In this case the value of β is point-identified (and can be estimated by the Hill estimator). Moreover, as in Property IV.3, the conditional survivor function of $\left(\frac{T_1^{\beta}}{t^{\beta}}-1,\frac{T_2^{\beta}}{t^{\beta}}-1\right)$ given $T_1>t,T_1>t$ converges also

to $H(x,y)$ when t goes to infinity. Indeed, the limit of this survivor function is equal to :

$$\lim_{t\to\infty}\frac{\mathbb{E}[e^{-\Lambda_1(t(1+x)^{\frac{1}{\beta}})U-\Lambda_2(t(1+y)^{\frac{1}{\beta}})V}]}{\mathbb{E}[e^{-\Lambda_1(t)U-\Lambda_2(t)V}]}=\lim_{t\to\infty}\frac{\mathcal{L}_{(U,V)}(\Lambda_1(t)\frac{\Lambda_1(t(1+x)^{\frac{1}{\beta}})}{\Lambda_1(t)},\Lambda_1(t)\frac{\Lambda_2(t(1+y)^{\frac{1}{\beta}})}{\Lambda_1(t)})}{\mathcal{L}_{(U,V)}(\Lambda_1(t),\Lambda_1(t)\frac{\Lambda_2(t)}{\Lambda_1(t)})}=\frac{\mathcal{L}_\mu(1+x,1+y)}{\mathcal{L}_\mu(1,1)}.$$

Equivalently, by a change of variable, we get :

$$\mathbb{P}\Big[\frac{T_1}{t}>1+z_1,\frac{T_2}{t}>1+z_2\mid T_1>t,T_2>t\Big]\to\frac{1}{c}\iint e^{-(1+z_1)^\beta u-(1+z_2)^\beta v}\mu(u,v)\mathrm{d}u\mathrm{d}v,$$

for all $z_1,z_2\geq 0$. In other words, (T_1,T_2) is regularly varying at infinity.

Proof of Lemma IV.4. We have : $\mathbb{P}[X>\tau,Y>\tau]=\dfrac{\iint e^{-u(1+\tau)-v(1+\tau)}\mu(u,v)\mathrm{d}u\mathrm{d}v}{\iint e^{-u-v}\mu(u,v)\mathrm{d}u\mathrm{d}v}=\dfrac{1}{(1+\tau)^\alpha}$. For part 2 of the lemma, we know, by Corollary C.1, that $p=\frac{\iota_1}{\iota_1+\iota_2}$, where $\iota_j, j=1,2$ are the two cause-specific intensity functions for duration variables X and Y, respectively. For instance,

$$\iota_1=-\frac{\partial}{\partial x}\log H(\tau,\tau)=\frac{\iint e^{-u(1+\tau)-v(1+\tau)}u\mu(u,v)\mathrm{d}u\mathrm{d}v}{\iint e^{-u(1+\tau)-v(1+\tau)}\mu(u,v)\mathrm{d}u\mathrm{d}v}.$$

Thus

$$\begin{aligned}p=\frac{\iota_1}{\iota_1+\iota_2}&=\frac{\iint e^{-u(1+\tau)-v(1+\tau)}u\mu(u,v)\mathrm{d}u\mathrm{d}v}{\iint e^{-u(1+\tau)-v(1+\tau)}(u+v)\mu(u,v)\mathrm{d}u\mathrm{d}v}\\&=\frac{(1+\tau)^{-\alpha-1}\iint e^{-u-v}u\mu(u,v)\mathrm{d}u\mathrm{d}v}{\frac{\mathrm{d}}{\mathrm{d}\tau}\iint e^{-u(1+\tau)-v(1+\tau)}\mu(u,v)\mathrm{d}u\mathrm{d}v}\\&=\frac{(1+\tau)^{-\alpha-1}\iint e^{-u-v}u\mu(u,v)\mathrm{d}u\mathrm{d}v}{\frac{\mathrm{d}}{\mathrm{d}\tau}(1+\tau)^{-\alpha}\iint e^{-u-v}\mu(u,v)\mathrm{d}u\mathrm{d}v}=\frac{\iint e^{-u-v}u\mu(u,v)\mathrm{d}u\mathrm{d}v}{\alpha\iint e^{-u-v}\mu(u,v)\mathrm{d}u\mathrm{d}v},\end{aligned}$$

by the homogeneity property of μ. □

Bibliographie

Abbring, J. and van den Berg, G. (2003a). The Identifiability of the Mixed Proportional Hazards Competing Risks Model. *Journal of the Royal Statistical Society : Series B*, 65(3) :701–710.

Abbring, J. and van den Berg, G. (2003b). The Nonparametric Identification of Treatment Effects in Duration Models. *Econometrica*, 71(5) :1491–1517.

Abbring, J. and van den Berg, G. (2007). The Unobserved Heterogeneity Distribution in Duration Analysis. *Biometrika*, 94(1) :87–99.

Amemiya, T. (1984). Tobit Models : a Survey. *Journal of Econometrics*, 24(1) :3–61.

Andersen, P. and Keiding, N. (2012). Interpretability and Importance of Functionals in Competing Risks and Multistate Models. *Statistics in Medicine*, 31(11-12) :1074–1088.

Bai, J., Collin-Dufresne, P., Goldstein, R., and Helwege, J. (2014). On Bounding Credit Event Risk Premia. *forthcoming Review of Financial Studies*.

Baker, M. and Melino, A. (2000). Duration Dependence and Nonparametric Heterogeneity : A Monte Carlo Study. *Journal of Econometrics*, 96(2) :357–393.

Bauer, D., Börger, M., and Ruß, J. (2010). On the Pricing of Longevity-Linked Securities. *Insurance : Mathematics and Economics*, 46(1) :139–149.

Benzoni, L., Collin-Dufresne, P., Goldstein, R., and Helwege, J. (2012). Modeling Credit Contagion via the Updating of Fragile Beliefs. *Federal Reserve Bank of Chicago Working Paper*.

Berg, C., Christensen, J. P. R., and Ressel, P. (1984). *Harmonic Analysis on Semigroups*. Springer-Verlag.

Brouhns, N., Denuit, M., and Vermunt, J. (2002). A Poisson Log-Bilinear Regression Approach to the Construction of Projected Life Tables. *Insurance : Mathematics and Economics*, 31(3) :373–393.

Cairns, A., Blake, D., and Dowd, K. (2006). A Two-Factor Model for Stochastic Mortality with Parameter Uncertainty : Theory and Calibration. *Journal of Risk and Insurance*, 73(4) :687–718.

Cairns, A., Blake, D., Dowd, K., Coughlan, G., and Epstein, D. (2009). A Quantitative Comparison of Stochastic Mortality Models Using Data from England & Wales and the United States. *North American Actuarial Journal*, 13(1) :1–35.

Carriere, J. (2000). Bivariate Survival Models for Coupled Lives. *Scandinavian Actuarial Journal*, (1) :17–32.

Charpentier, A. and Segers, J. (2009). Tails of Multivariate Archimedean Copulas. *Journal of Multivariate Analysis*, 100(7) :1521–1537.

Chen, R. and Scott, L. (1993). Maximum Likelihood Estimation of a Multi-Factor Equilibrium Model of the Term Structure of Interest Rates. *Journal of Fixed Income*, (3) :14–31.

Chen, X., Chernozhukov, V., Lee, S., and Newey, W. K. (2014). Local Identification of Nonparametric and Semiparametric Models. *Econometrica*, 82(2) :785–809.

Clayton, D. (1978). A Model for Association in Bivariate Life Tables and its Application in Epidemiological Studies of Familial Tendency in Chronic Disease Incidence. *Biometrika*, 65(1) :141–151.

Cox, J. C., Ingersoll Jr, J. E., and Ross, S. A. (1985). A Theory of the Term Structure of Interest Rates. *Econometrica*, 53(2) :385–407.

Czado, C., Delwarde, A., and Denuit, M. (2005). Bayesian Poisson Log-Bilinear Mortality Projections. *Insurance : Mathematics and Economics*, 36(3) :260–284.

de Haan, L. and Omey, E. (1984). Integrals and Derivatives of Regularly Varying Functions in $\mathbb{R}^d$ and Domains of Attraction of Stable Distributions II. *Stochastic Processes and their Applications*, 16(2) :157–170.

de Haan, L., Omey, E., and Resnick, S. (1984). Domains of Attraction and Regular Variation in $\mathbb{R}^d$. *Journal of Multivariate Analysis*, 14(1) :17–33.

de Haan, L. and Resnick, S. I. (1979). Derivatives of Regularly Varying Functions in $\mathbb{R}^d$ and Domains of Attraction of Stable Distributions. *Stochastic Processes and their Applications*, 8(3) :349–355.

Deng, Y., Quigley, J., and Order, R. (2000). Mortgage Terminations, Heterogeneity and the Exercise of Mortgage Options. *Econometrica*, 68(2) :275–307.

Denuit, M. and Cornet, A. (1999). Multilife Premium Calculation with Dependent Future Lifetimes. *Journal of Actuarial Practice*, 7 :147–171.

Denuit, M., Dhaene, J., Le Bailly de Tilleghem, C., Teghem, S., et al. (2001). Measuring the Impact of Dependence Among Insured Life Lengths. *Belgian Actuarial Bulletin*, 1(1) :18–39.

Dienst, H.-R. (1972). Gedanken zum Konjunkturrisiko in der Berufsunfähigkeits-Versicherung. *Blätter der DGVFM*, 10(4) :521–531.

Dionne, G. (1982). Moral Hazard and State-Dependent Utility Function. *Journal of Risk and Insurance*, pages 405–422.

Duffie, D., Eckner, A., Horel, G., and Saita, L. (2009). Frailty Correlated Default. *Journal of Finance*, 64(5) :2089–2123.

Duffie, D., Pan, J., and Singleton, K. (2000). Transform Analysis and Asset Pricing for Affine Jump-Diffusions. *Econometrica*, 68(6) :1343–1376.

Elbers, C. and Ridder, G. (1982). True and Spurious Duration Dependence : The Identifiability of the Proportional Hazard Model. *Review of Economic Studies*, 49(3) :403–409.

Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*, volume 33. Springer.

Esscher, F. (1932). On the Probability Function in the Collective Theory of Risk. *Skandinavisk Aktuarietidskift*, 15 :175–195.

Feller, W. (2008). *An Introduction to Probability Theory and its Applications, Second Edition*, volume 2. John Wiley & Sons.

Fine, J., Jiang, H., and Chappell, R. (2001). On Semi-Competing Risks Data. *Biometrika*, 88(4) :907–919.

Finkelstein, A. and McGarry, K. (2006). Multiple Dimensions of Private Information : Evidence From the Long-Term Care Insurance Market. *American Economic Review*, 96(4) :938–958.

Frees, E., Carriere, J., and Valdez, E. (1996). Annuity Valuation with Dependent Mortality. *Journal of Risk and Insurance*, 63(2) :229–261.

Freund, J. (1961). A Bivariate Extension of the Exponential Distribution. *Journal of the American Statistical Association*, 56(296) :971–977.

Gagliardini, P. and Gouriéroux, C. (2014). *Granularity Theory with Applications to Finance and Insurance*. forthcoming, Cambridge University Press.

Gagliardini, P., Gouriéroux, C., and Monfort, A. (2012). Microinformation, Nonlinear Filtering, and Granularity. *Journal of Financial Econometrics*, 10(1) :1–53.

Genest, C. and MacKay, J. (1986). Copules Archimédiennes et Familles de Lois Bidimensionnelles dont les Marges sont Données. *Canadian Journal of Statistics*, 14(2) :145–159.

Gerber, H. and Shiu, E. (1994). Option Pricing by Esscher Transforms. *Transactions of the Society of Actuaries*, 46 :99–140.

Gouriéroux, C. and Jasiak, J. (2006). Autoregressive Gamma Processes. *Journal of Forecasting*, 25(2) :129–152.

Gouriéroux, C. and Lu, Y. (2013). Love and Death : A Freund Model with Frailty. *CREST DP*.

Gourieroux, C., Monfort, A., and Renne, J.-P. (2014). Pricing Default Events : Surprise, Exogeneity and Contagion. *Journal of Econometrics*, 182(2) :397–411.

Gruber, J. and Wise, D. (1999). Social Security Programs and Retirement around the World. *Research in Labor Economics*, 18(1) :40.

Haghani, S. (2014). Modelling Hedge Fund Lifetimes : A Dependent Competing Risks Framework with Latent Exit Types. *Journal of Empirical Finance*, 28 :291–320.

Hammersley, J. M. and Clifford, P. (1968). *Markov Fields on Finite Graphs and Lattices*. Oxford University DP.

Han, A. and Hausman, J. A. (1990). Flexible Parametric Estimation of Duration and Competing Risk Models. *Journal of Applied Econometrics*, 5(1) :1–28.

Heckman, J. and Robb, R. (1985). Using Longitudinal Data to Estimate Age, Period and Cohort Effects in Earnings Equations. In *Cohort Analysis in Social Research*, pages 137–150. Springer.

Heckman, J. and Singer, B. (1984). A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data. *Econometrica*, 52(2) :271–320.

Honoré, B. and Lleras-Muney, A. (2006). Bounds in Competing Risks Models and the War on Cancer. *Econometrica*, 74(6) :1675–1698.

Honoré, B. E. (1993). Identification Results for Duration Models with Multiple Spells. *Review of Economic Studies*, 60(1) :241–246.

Horowitz, J. L. (1999). Semiparametric Estimation of a Proportional Hazard Model with Unobserved Heterogeneity. *Econometrica*, 67(5) :1001–1028.

Hougaard, P. (2000). *Analysis of Multivariate Survival Data*, volume 564. Springer New York.

Hsieh, J.-J., Wang, W., and Adam Ding, A. (2008). Regression Analysis Based on Semicompeting Risks Data. *Journal of the Royal Statistical Society : Series B*, 70(1) :3–20.

Iachine, I. (2004). Identifiability of Bivariate Frailty Models. *University of Southern Denmark Working Paper.*

Imai, K. and Soneji, S. (2007). On the Estimation of Disability-Free Life Expectancy : Sullivan's Method and its Extension. *Journal of the American Statistical Association*, 102(480) :1199–1211.

Ishwaran, H. (1996). Identifiability and Rates of Estimation for Scale Parameters in Location Mixture Models. *Annals of Statistics*, 24(4) :1560–1571.

Jagger, C. and Sutton, C. (1991). Death After Marital Bereavement-Is the Risk Increased ? *Statistics in Medicine*, 10(3) :395–404.

Jarrow, R. and Yu, F. (2001). Counterparty Risk and the Pricing of Defaultable Securities. *The Journal of Finance*, 56(5) :1765–1799.

Ji, M., Hardy, M., and Li, S. (2011). Markovian Approaches to Joint-Life Mortality. *North American Actuarial Journal*, 15(3) :357–376.

Juri, A. and Wüthrich, M. V. (2002). Copula Convergence Theorems for Tail Events. *Insurance : Mathematics and Economics*, 30(3) :405–420.

Karni, E. (1983). Risk Aversion for State-Dependent Utility Functions : Measurement and Applications. *International Economic Review*, 24(3) :637–647.

Keiding, N. (1991). Age-Specific Incidence and Prevalence : a Statistical Perspective. *Journal of the Royal Statistical Society : Series A*, 154(3) :371–412.

Keilman, N., Pham, D. Q., and Hetland, A. (2002). Why Population Forecasts Should be Probabilistic–Illustrated by the Case of Norway. *Demographic Research*, 6(15) :409–454.

Kessler, D. (2008). The Long-Term Care Insurance Market. *The Geneva Papers on Risk and Insurance-Issues and Practice*, 33(1) :33–40.

Kuang, D., Nielsen, B., and Nielsen, J. (2008). Identification of the Age-period-Cohort Model and the Extended Chain-Ladder Model. *Biometrika*, 95(4) :979–986.

Lafortune, G. and Balestat, G. (2007). Trends in Severe Disability Among Elderly People : Assessing the Evidence in 12 OECD Countries and the Future Implications. *OECD Health Working Paper*.

Lancaster, T. (1979). Econometric Methods for the Duration of Unemployment. *Econometrica*, 47(4) :939–956.

Ledford, A. W. and Tawn, J. A. (1996). Statistics for Near Independence in Multivariate Extreme Values. *Biometrika*, 83(1) :169–187.

Lee, R. and Carter, L. (1992). Modeling and Forecasting US Mortality. *Journal of the American Statistical Association*, 87(419) :659–671.

Levantesi, S. and Menzietti, M. (2012). Managing Longevity and Disability Risks in Life Annuities with Long Term Care. *Insurance : Mathematics and Economics*, 50(3) :391–401.

Li, J. S.-H. and Hardy, M. R. (2011). Measuring Basis Risk in Longevity Hedges. *North American Actuarial Journal*, 15(2) :177–200.

Luciano, E., Spreeuw, J., and Vigna, E. (2008). Modelling Stochastic Mortality for Dependent Lives. *Insurance : Mathematics and Economics*, 43(2) :234–244.

Luciano, E., Spreeuw, J., and Vigna, E. (2010). Cross-Generational Comparison of Stochastic Mortality of Coupled Lives. *University of Torino Working Paper*.

Majer, I. M., Stevens, R., Nusselder, W. J., Mackenbach, J. P., and van Baal, P. H. (2013). Modeling and Forecasting Health Expectancy : Theoretical Framework and Application. *Demography*, 50(2) :673–697.

Mammen, E., Nielsen, J., and Fitzenberger, B. (2011). Generalized Linear Time Series Regression. *Biometrika*, 98(4) :1007–1014.

Manton, K. G., Stallard, E., and Corder, L. (1998). Economic Effects of Reducing Disability. *The American Economic Review*, 88(2) :101–105.

Marshall, A. W. and Olkin, I. (1988). Families of Multivariate Distributions. *Journal of the American Statistical Association*, 83(403) :834–841.

McNeil, A. and Nešlehová, J. (2009). Multivariate Archimedean Copulas, d-Monotone Functions and l_1-Norm Symmetric Distributions. *Annals of Statistics*, 37(5B) :3059–3097.

McWhinnie, J. R. (1980). Disability Assessment in Population Surveys : Results of the OECD Common Development Effort. *Revue d'Épidémiologie et de Santé Publique*, 29(4) :413–419.

Murtaugh, C. M., Spillman, B. C., and Warshawsky, M. J. (2001). In Sickness and in Health : An Annuity Approach to Financing Long-Term Care and Retirement Income. *Journal of Risk and Insurance*, 68(2) :225–254.

Nelsen, R. (1999). *An Introduction to Copulas*. Springer Verlag, New-York.

Nihtilä, E. and Martikainen, P. (2008). Institutionalization of Older Adults After the Death of a Spouse. *American Journal of Public Health*, 98(7) :1228.

Norberg, R. (1989). Actuarial Analysis of Dependent Lives. *Bulletin of the Swiss Association of Actuaries*, (2) :243–254.

Oakes, D. (1989). Bivariate Survival Models Induced by Frailties. *Journal of the American Statistical Association*, 84(406) :487–493.

Oeppen, J. and Vaupel, J. W. (2002). Broken Limits to Life Expectancy. *Science*, 296(5570) :1029–1031.

Parkes, M., Benjamin, B., and Fitzgerald, R. (1969). Broken Heart : A Statistical Study of Increased Mortality Among Widowers. *British Medical Journal*, 1(5646) :740–743.

Pedroza, C. (2006). A Bayesian Forecasting Model : Predicting US Male Mortality. *Biostatistics*, 7(4) :530–550.

Prentice, R. L., Kalbfleisch, J. D., Peterson Jr, A. V., Flournoy, N., Farewell, V., and Breslow, N. (1978). The Analysis of Failure Times in the Presence of Competing Risks. *Biometrics*, 34(4) :541–554.

Renshaw, A. and Haberman, S. (2003). Lee-Carter Mortality Forecasting with Age-Specific Enhancement. *Insurance : Mathematics and Economics*, 33(2) :255–272.

Renshaw, A. and Haberman, S. (2006). A Cohort-Based Extension to the Lee-Carter Model for Mortality Reduction Factors. *Insurance : Mathematics and Economics*, 38(3) :556–570.

Resnick, S. I. (2007). *Heavy-Tail Phenomena : Probabilistic and Statistical Modeling*. Springer.

Revuz, D. and Yor, M. (1999). *Continuous Martingales and Brownian Motion*, volume 293. Springer.

Rice, D. P. (1989). Health and Long-Term Care for the Aged. *The American Economic Review*, 79(2) :343–348.

Ridder, G. (1990). The Non-Parametric Identification of Generalized Accelerated Failure-Time Models. *Review of Economic Studies*, 57(2) :167–181.

Shemyakin, A. and Youn, H. (2006). Copula Models of Joint Last Survivor Analysis. *Applied Stochastic Models in Business and Industry*, 22(2) :211–224.

Sklar, M. (1959). Fonctions de Répartition à n Dimensions et Leurs Marges. *Publications de l'Institut de Statistique de l'Université de Paris*, (8) :229–231.

Spillman, B. C. and Lubitz, J. (2002). New Estimates of Lifetime Nursing Home Use : Have Patterns of Use Changed ? *Medical Care*, 40(10) :965–975.

Spreeuw, J. (2006). Types of Dependence and Time-Dependent Association Between Two Lifetimes in Single Parameter Copula Models. *Scandinavian Actuarial Journal*, (5) :286–309.

Spreeuw, J. and Owadally, I. (2013). Investigating the Broken-Heart Effect : a Model for Short-Term Dependence Between the Remaining Lifetimes of Joint Lives. *Annals of Actuarial Science*, 7(2) :236–257.

Spreeuw, J. and Wang, X. (2008). Modelling the Short-term Dependence Between Two Remaining Lifetimes. *Cass Business School Discussion Paper*.

Tosch, T. and Holmes, P. (1980). A Bivariate Failure Model. *Journal of the American Statistical Association*, 75(370) :415–417.

Tsiatis, A. (1975). A Nonidentifiability Aspect of the Problem of Competing Risks. *Proceedings of the National Academy of Sciences*, 72(1) :20–22.

Vaupel, J., Manton, K., and Stallard, E. (1979). The Impact of Heterogeneity in Individual Frailty on the Dynamics of Mortality. *Demography*, 16(3) :439–454.

Wills, S. and Sherris, M. (2010). Securitization, Structuring and Pricing of Longevity Risk. *Insurance : Mathematics and Economics*, 46(1) :173–185.

Wouterse, B., Huisman, M., Meijboom, B. R., Deeg, D. J., and Polder, J. J. (2013). Modeling the Relationship Between Health and Health Care Expenditures Using a Latent Markov Model. *Journal of Health Economics*, 32(2) :423–439.

Xu, J., Kalbfleisch, J. D., and Tai, B. (2010). Statistical Analysis of Illness–Death Processes and Semicompeting Risks Data. *Biometrics*, 66(3) :716–725.

Xue, X. and Brookmeyer, R. (1996). Bivariate Frailty Model for the Analysis of Multivariate Survival Time. *Lifetime Data Analysis*, 2(3) :277–289.

Yashin, A. I., Vaupel, J. W., and Iachine, I. A. (1995). Correlated Individual Frailty : an Advantageous Approach to Survival Analysis of Bivariate Data. *Mathematical Population Studies*, 5(2) :145–159.

Youn, H. and Shemyakin, A. (1999). Statistical Aspects of Joint Life Insurance Pricing. *Proceedings of the Business and Statistics Section of the American Statistical Association*.

Youn, H. and Shemyakin, A. (2001). Pricing Practices for Joint Last Survivor Insurance. *Actuarial Research Clearing House*, 1(2) :3.

Youn, H., Shemyakin, A., and Herman, E. (2002). A Re-Examination of the Joint Mortality Functions. *North American Actuarial Journal*, 6(1) :166–170.

Zheng, M. and Klein, J. (1995). Estimates of Marginal Survival for Dependent Competing Risks Based on an Assumed Copula. *Biometrika*, 82(1) :127–138.

Zocher, M. (2006). Multivariate Counting Processes. *TU Dresden Working Paper*.